

Estimation of the Impact of Sampling Errors in the VOS Observations on Air–Sea Fluxes. Part I: Uncertainties in Climate Means

SERGEY GULEV

P. P. Shirshov Institute of Oceanology, Moscow, Russia, and IFM-GEOMAR, Kiel, Germany

THOMAS JUNG

ECMWF, Reading, United Kingdom

EBERHARD RUPRECHT

IFM-GEOMAR, Kiel, Germany

(Manuscript received 28 July 2005, in final form 2 June 2006)

ABSTRACT

Sampling uncertainties in the voluntary observing ship (VOS)-based global ocean–atmosphere flux fields were estimated using the NCEP–NCAR reanalysis and ECMWF 40-yr Re-Analysis (ERA-40) as well as seasonal forecasts without data assimilation. Air–sea fluxes were computed from 6-hourly reanalyzed individual variables using state-of-the-art bulk formulas. Individual variables and computed fluxes were subsampled to simulate VOS-like sampling density. Random simulation of the number of VOS observations and simulation of the number of observations with contemporaneous sampling allowed for estimation of random and total sampling uncertainties respectively. Although reanalyses are dependent on VOS, constituting an important part of data assimilation input, it is assumed that the reanalysis fields adequately reproduce synoptic variability at the sea surface. Sampling errors were quantified by comparison of the regularly sampled (i.e., 6 hourly) and subsampled monthly fields of surface variables and fluxes. In poorly sampled regions random sampling errors amount to 2.5°–3°C for air temperature, 3 m s^{−1} for the wind speed, 2–2.5 g kg^{−1} for specific humidity, and 15%–20% of the total cloud cover. The highest random sampling errors in surface fluxes were found for the sensible and latent heat flux and range from 30 to 80 W m^{−2}. Total sampling errors in poorly sampled areas may be higher than random ones by 60%. In poorly sampled subpolar latitudes of the Northern Hemisphere and throughout much of the Southern Ocean the total sampling uncertainty in the net heat flux can amount to 80–100 W m^{−2}. The highest values of the uncertainties associated with the interpolation/extrapolation into unsampled grid boxes are found in subpolar latitudes of both hemispheres for the turbulent fluxes, where they can be comparable with the sampling errors. Simple dependencies of the sampling errors on the number of samples and the magnitude of synoptic variability were derived. Sampling errors estimated from different reanalyses and from seasonal forecasts yield qualitatively comparable spatial patterns, in which the actual values of uncertainties are controlled by the magnitudes of synoptic variability. Finally, estimates of sampling uncertainties are compared with the other errors in air–sea fluxes and the reliability of the estimates obtained is discussed.

1. Introduction

Accurate knowledge of air–sea flux fields is required for global and regional budget studies, forcing ocean general circulation models and analysis of climate variability. Global and regional surface flux fields are pres-

ently available from atmospheric modeling, remote sensing, and from the computations of fluxes using voluntary observing ship (VOS) data. Complete space–time coverage of fluxes and related fields are now available from numerical weather prediction (NWP) models run with fixed data assimilation systems and historical observations widely referred to as “reanalyses” (Kalnay et al. 1996; Kistler et al. 2001; Gibson et al. 1999; Upala et al. 2005) and operational analyses. However, their validation shows significant biases in different flux components (Josey 2001; Moore and Renfrew 2002;

Corresponding author address: Sergey Gulev, P. P. Shirshov Institute of Oceanology, RAS, 36 Nakhimovsky Ave., 117851 Moscow, Russia.
E-mail: gul@sail.msk.ru

Rouault et al. 2003; Smith et al. 2001). Satellite flux products (Schulz et al. 1997; The WGASF Group 2000; Bentamy et al. 2003; Chou et al. 2003; Smith et al. 2004) also exhibit regional and global biases, which may locally amount to as much as $30\text{--}50\text{ W m}^{-2}$ (e.g., Kubota et al. 2003; The WGASF Group 2000). Moreover, satellite flux products are still too short in time to be effectively used for climate variability studies.

Because of the shortcomings of NWP and satellite flux products, the flux fields based on VOS data (da Silva et al. 1994; Josey et al. 1999; Lindau 2000) remain important for sea–air interaction studies on global and regional scales. The last update of the International Comprehensive Ocean–Atmosphere Data Set (ICOADS) formally provides marine meteorological data starting from 1784 (Worley et al. 2005), making it an extremely attractive dataset for climate-related studies. However, VOS-based surface fluxes suffer from the uncertainties inherent in ship observations, parameterizations, and the methodology used to produce surface flux fields. The effect of different parameterizations on air–sea flux climatologies (Blanc 1985, 1986; Gulev 1995; Zeng et al. 1998; Brunke et al. 2003; Bignami et al. 1995; Josey et al. 1997; Dobson and Smith 1988) results in typical uncertainties of $10\text{--}15\text{ W m}^{-2}$ for sensible heat and $30\text{--}40\text{ W m}^{-2}$ for latent heat flux. Uncertainties in shortwave (SW) and longwave (LW) radiation vary within $10\text{--}25\text{ W m}^{-2}$ and $5\text{--}15\text{ W m}^{-2}$, respectively. Uncertainties in climatological fluxes due to corrections of the routinely measured and estimated variables (Kent et al. 1993a,b) locally amount to $\pm 15\text{ W m}^{-2}$ with the largest contribution from the corrections of air temperature and humidity (Josey et al. 1999). However, the completeness of the corrections applied is still questionable because of the lack of a reference standard.

Besides the above-mentioned uncertainties, VOS-based flux products suffer from poor and inhomogeneous sampling. ICOADS sampling varies regionally, ranging from zero to several thousand samples per $2^\circ \times 2^\circ$ box per calendar month. Moreover, in many areas there are time-dependent changes in the sampling density, associated with historical changes in the ship routes. For the tropical Pacific, Lo and McBean (1978), Weare and Strub (1981), and Weare (1989) reported small sampling errors in radiation fluxes and errors of several tens of Watts per meter squared in turbulent fluxes. However, these conclusions may not be necessarily valid for the midlatitudes and for highly variable parameters (Legler 1991). For selected North Atlantic locations, Cayan (1992) found that sampling errors in 5° monthly averaged turbulent fluxes sharply grow when the monthly number of reports becomes less than 10. The most advanced analysis of sampling errors was per-

formed for sea surface temperature (SST) (Kaplan et al. 2003; Smith and Reynolds 2004). Brohan et al. (2003) and Kent and Berry (2005) performed global analysis of errors in individual variables, including also estimation of sampling errors. These estimates were obtained using statistical analysis of the VOS data. However, a comprehensive analysis of sampling errors in VOS fluxes requires an alternative source of surface data, characterized by adequate sampling.

In this respect, the reanalyses' surface variables and fluxes are characterized by formally regular and homogeneous "sampling" over the global ocean. However, the reanalysis surface fields can still be affected by sampling problems. In well-sampled regions the reanalysis fields are strongly constrained by the observations and in poorly sampled regions they are dominated by the first guess and, hence, the atmospheric model used. Moreover, reanalysis systems assimilate relatively well observed parameters like terrestrial network input and radiosonde observations, which are not part of the ICOADS dataset. Bengtsson et al. (2004) analyzed the impact of different data on the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr Re-Analysis (ERA-40). They found that the surface observational system alone is insufficient to produce reliable atmospheric fields due to poor representation of vortices whose reliable representation depends crucially on other sources of observations. Sterl (2001) has shown that more reliable variability patterns can be derived from reanalysis compared to VOS data, attributing the disagreement to the gap-filling algorithms in poorly sampled areas.

Limited subsets of NWP data were used for assessing sampling errors in the midtropospheric flow (Kidson and Trenberth 1988) and ICOADS sea level pressure (SLP) (Chang 2005). Use of model data for the estimation of sampling impact on the atmospheric humidity and thermal energy transports and temperature–humidity relationships was first analyzed in the pioneering work of Oort (1978) and also studies by Bauer et al. (2002). However, the impact of sampling on the air–sea fluxes is much more complicated than that on individual variables. In this study, which aims to assess the role of sampling errors in the computation of surface energy fluxes from the VOS data, we will use the VOS data from the latest ICOADS updates along with surface variables and fluxes from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) and ERA-40 reanalyses along with NWP simulations without data assimilation. Our strategy is to simulate VOS-like sampling in the reanalysis data and attribute the differences between the subsampled and the original NWP

fields (the “truth”) to sampling errors. By using different NWP products we will be able to understand how sensitive our error estimates are to the choice of reanalyses. We will first analyze sampling uncertainties in individual variables and then show how these errors can affect the flux estimates.

2. Data

For the estimation of sampling uncertainties over the World Ocean we used the recent update of ICOADS Long Marine Records Format, version 6.0 (LMRF-6.0) for the period 1948–2002, which was specially reprocessed in support of the NCEP–NCAR reanalysis project (Woodruff et al. 1998; NCAR 1999; Worley et al. 2005). The ICOADS collection includes several types of platforms (merchant and research vessels, moored and drifting buoys, etc.). Being focused on the actual number of the observations that can be used for the flux computations, we used all platforms, except for the drifting buoys whose number increased 10 times from the 1980s, and which currently contribute $\approx 50\%$ of all reports. Most drifting buoys report SST only, with some reporting SLP and just a few reporting wind, and thus cannot be used for the flux computations. We will use the term “VOS” for identifying all ICOADS platforms, excluding drifting buoys.

From the NCEP–NCAR reanalysis (Kalnay et al. 1996; Kistler et al. 2001), 6-hourly individual variables (10-m wind speed, SST, 2-m air temperature and humidity, SLP, total cloud cover) at $1.875^\circ \times \approx 1.9^\circ$ resolution (Gaussian grid) were used for the same 55-yr period as the VOS data. In addition to the NCEP–NCAR reanalysis we also used ERA-40 reanalysis data for the period 1992–2001. The ERA-40 (1958–2001) system (Uppala et al. 2005) uses a higher-resolution AGCM (T_L159) and assimilated Special Sensor Microwave Imager (SSM/I) radiances as well as *European Remote Sensing Satellite-1 and -2 (ERS-1/2)* data in more recent years. To quantify the impact of actual data on the estimates of sampling errors obtained from ERA-40 we also used winter seasonal integrations (started on 1 October) for the same 10-yr period (1992–2001). Although these were carried out with a more recent version of the ECMWF model (cycle 28r3), the same resolution (T_L159 and 60 levels in the vertical) and SST as in the ERA-40 system were used. We interpret these integrations as a “data assimilation experiment” without the use of any observations (i.e., the resulting “analysis fields” are effectively a sequence of first guesses). The output of both ERA-40 reanalysis and ECMWF forecasts (ECF hereafter) was at 6-hourly temporal and 2.5° spatial resolution. For the further estimation of sampling errors all fields were interpo-

lated onto a $2^\circ \times 2^\circ$ grid by the method of local procedures of Akima (1970).

Finally, we use a regional subset of high-quality meteorological data collected in the Northwest Atlantic during the period 1980–92 under the SECTIONS (RAZREZY in Russian) program (Gulev 1994, 1999). For selected seasons these data have very high sampling density and just less than 20% of them were incorporated into ICOADS. These data provided quasi-independent estimates of sampling uncertainties using alternative high-quality data collection.

3. Methods and strategy

a. Characteristics of sampling uncertainties in VOS flux fields

Sampling uncertainties in monthly VOS-based air-sea fluxes result from the inadequate representation of the synoptic variability by the irregularly sampled VOS observations. Many ICOADS reports contain certain individual variables (e.g., air temperature or wind), missing others (e.g., humidity or cloud cover). This implies different sampling uncertainties for different variables and affects the ICOADS Monthly Summary Trimmed Groups (MSTG), that is, 1° or 2° averages of basic variables (Woodruff et al. 1998). However, the flux computation requires exclusively those reports providing all variables needed for the proper application of bulk formulas. Figure 1a shows the average monthly mean number of air temperature (the best sampled variable) reports per $2^\circ \times 2^\circ$ box for the period 1948–2002. Areas of substantial undersampling for air temperature and other individual parameters (not shown) are found in the Southern Ocean and Northern Hemisphere subpolar regions. This becomes especially true if only those reports that provide all parameters needed for the flux computations are shown (Fig. 1b). Their number is 2–15 times smaller than that for individual variables, showing south of 40°S for most $2^\circ \times 2^\circ$ boxes less than 0.1 reports per month.

Figure 1c shows the temporal distribution of the total number of ICOADS reports (excluding drifting buoys) containing selected variables and all flux-related variables. The total number of air temperature and wind speed reports generally increased during 1948–2002. Until the late 1970s, the total number of cloud cover reports was comparable to those for air temperature and wind speed but started to decrease quite rapidly in the mid-1980s. The total number of reports containing all flux-related variables is substantially smaller throughout the whole period. The number of reports containing all flux-related variables did not change much during the last three decades, yet in relative terms

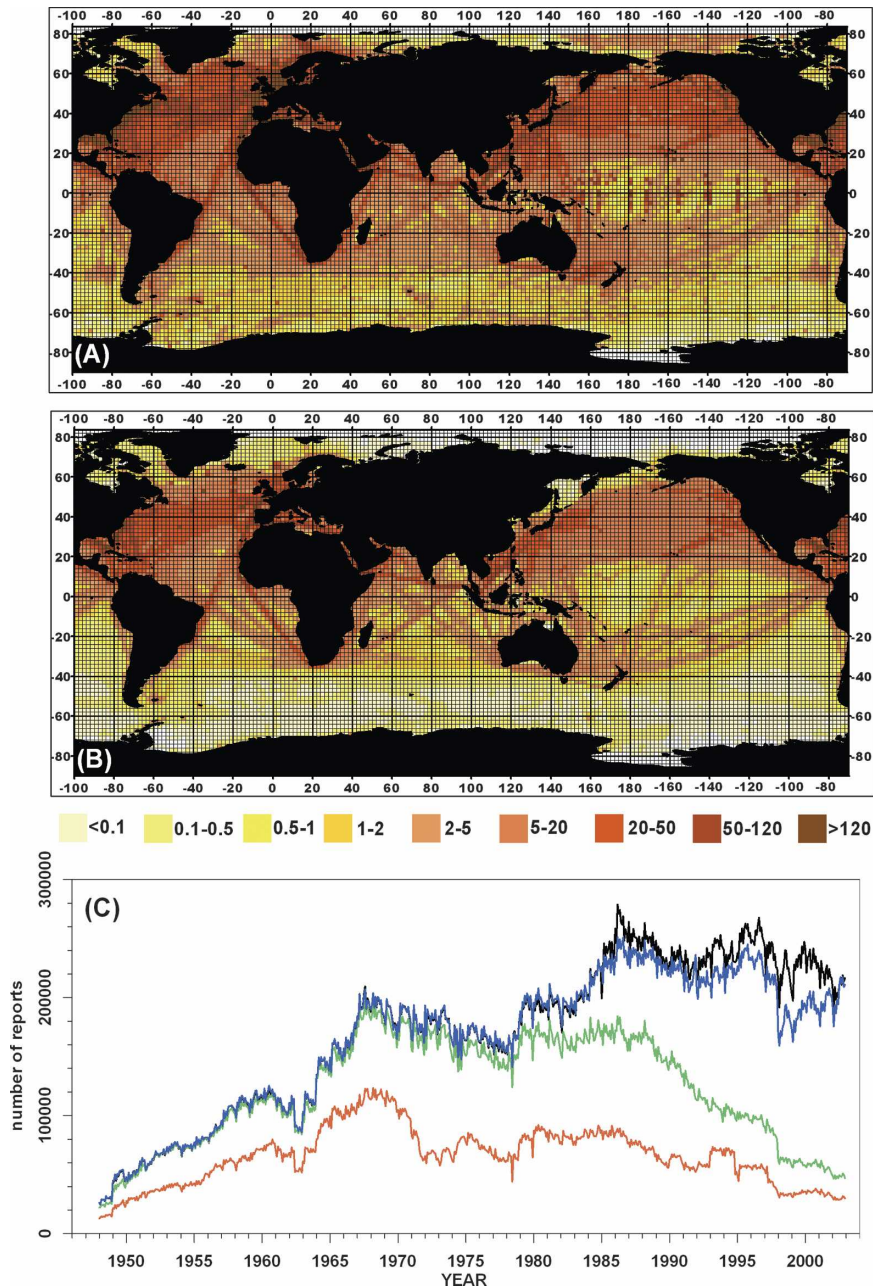


FIG. 1. (a) Distribution of the average monthly number of air temperature observations in ICOADS, as well as (b) the average monthly number of reports, containing all individual flux-related variables. The average was taken over 660 calendar months of the period 1948–2002. (c) Time series of the total number of reports for air temperature (black), wind speed (blue), total cloud cover (green), and those containing all flux-related variables (red).

their number actually decreased substantially. After the early 1970s the percentage of reports with all flux-related variables decreased to 10%–25% of the total number of reports in comparison to 40%–60% during 1948–70. Thus, for the surface flux computations the 1960s are still the best-sampled decade. Besides the

moored buoys, whose installation has started in the 1980s, this can be explained by the extended installation of automatic packages, providing measurements of wind speed, air temperature, and SST onboard marine carriers starting from the early 1970s. Having such a package onboard, marine officers typically do not prac-

tice the observations of the variables not measured automatically (e.g., cloud cover).

In summary, Fig. 1 implies that sampling uncertainties in flux-related variables and computed fluxes are spatially dependent, being higher in the Southern Ocean than in the Northern Hemisphere midlatitudes. Moreover, different variables are influenced to a different extent by inadequate sampling. Sea–air fluxes are always more strongly influenced by sampling problems than individual variables are. Finally, sampling uncertainties are time dependent due to temporal changes in the number of observations. This is particularly true for the locations of the Ocean Weather Stations (OWS) in the North Atlantic and North Pacific, where most observations were terminated in the early 1970s.

b. Sea–air flux computations from the NWP surface variables

The NWP products (NCEP–NCAR, ERA-40, and ECF) used in this study provide the turbulent and radiative fluxes computed by the operational model at the sea surface. However, the use of different parameterization schemes at different NWP centers may introduce additional uncertainty in the comparative assessment. To estimate the pure influence of sampling on VOS fluxes, here we use the individual variables from different NWP products and recomputed the fluxes for every 6-hourly step for the periods covered by the different products.

Sensible (Q_h) and latent (Q_e) heat fluxes were estimated using the Coupled Ocean/Atmosphere Response Experiment (COARE)-3.0 algorithm (Fairall et al. 2003), which is considered to be the least biased among the available surface flux schemes (Brunke et al. 2003). The NCEP–NCAR, ERA-40, and ECF air temperature and humidity (given at 2-m height) were adjusted to 10-m height, and surface wind speed (reported at 10 m) was adjusted to neutral stability using the flux-profile relationships from Fairall et al. (2003). The LW radiation was computed using the parameterization of Clark et al. (1974), used in Southampton Oceanography Centre (SOC) climatology (Josey et al. 1999). The different algorithms show quite a large range of estimates for turbulent heat fluxes (Blanc 1985; Kent and Taylor 1995; Hasse and Smith 1997; The WGASF Group 2000; Brunke et al. 2003) and for LW radiation (Fung et al. 1984; Bignami et al. 1995; Josey et al. 1997). Thus, quantitatively different results might be obtained for different schemes. However, we are focusing on the sampling impact on the flux fields. Qualitative similarity of the turbulent and LW schemes implies that our main conclusions will hold for most schemes.

For a proper computation of the SW radiation four observations per day (0000, 0600, 1200, and 1800 UTC) are not enough for an adequate description of the diurnal cycle, and as a result daily averaged SW fluxes will be biased. The SW flux diagnosed by the reanalyses is not a prognostic variable, but a 6-hourly *average* for the period of the reference time plus 6 h and thus accounts properly for the diurnal cycle. To apply the so-called hourly models designed for instantaneous measurements we computed the SW radiation, changing for each report the solar altitude around the clock with 30-min step for the midmonth date and the latitude varying within a corresponding 2° latitudinal belt, assuming the cloud cover and temperature to be constant (Gulev 2006, manuscript submitted to *J. Atmos. Oceanic Technol.*). In some regions diurnal variations in cloudiness and temperature may affect the results. This bias, however, is more important for climate means than for estimated sampling uncertainties. For the computations of instantaneous SW values we used the model of Malevsky et al. (1992), which has been employed for the construction of the Institut für Meereskunde (IFM) climatology (Lindau 2000). Its validation against observational data (Niekamp 1992; Gulev 1995) shows that it gives the smallest biases in comparison to the parameterizations of Reed (1977), Dobson and Smith (1988), and Lind et al. (1984). Instead of the astronomically driven Payne's (1972) albedo, traditionally used in VOS computations, we applied the Girdiuk et al. (1985) albedo, which depends also on the cloud cover and gives more realistic values in high latitudes.

c. Simulation of the VOS sampling density in the NWP flux-related variables and fluxes

The regularly sampled flux-related variables and recomputed fluxes from the NWP products were subsampled to match the VOS sampling density. In different regions ICOADS provides from zero to several thousand samples per months in 2° × 2° boxes, whereas NCEP–NCAR, ERA-40, and ECF provide always 120 samples for a 30-day “month.” The limit of 120 samples per month is exceeded in VOS only over midlatitudinal areas of the Northern Hemisphere (Fig. 1). To simulate the VOS-like sampling for the NWP flux fields, 6-hourly data were matched to the dates and UTC hours of the ICOADS reports. If the ICOADS reports were not available at “standard” times (0000, 0600, 1200, and 1800 UTC), the adjacent snapshots were interpolated to the exact time of the VOS sample. In cases where more than one ICOADS report was available at one instant, the NWP fluxes were repeatedly

used to simulate the oversampling of the VOS data. A similar procedure was used by Gulev et al. (2003b) for estimating sampling uncertainties in wave parameters. Da Silva et al. (1994) used a simplified version of this procedure for the comparison of University of Wisconsin–Milwaukee (UWM) and 12-hourly ECMWF operational fields. This procedure was applied to the turbulent fluxes, LW radiation, and the SW radiation recomputed for the “rotated around the clock astronomy” and midmonth date (Gulev 2006, manuscript submitted to *J. Atmos. Oceanic Technol.*).

After the procedure of subsampling, the computed fluxes and individual variables were averaged within $2^\circ \times 2^\circ$ boxes. This procedure leaves gaps for each month in the unsampled $2^\circ \times 2^\circ$ boxes. Application of different interpolation methods for gap filling gives yet another source of uncertainties of VOS products (see, e.g., Sterl 2001, 2004). In this study we applied the modified method of local procedures (Akima 1970) in order to produce flux fields on a regular grid, which was used for the development of the wave climatology (Gulev et al. 2003b). The idea behind this method is close to the kriging technique, but is much more computationally effective. Spatial smoothing was done using a two-dimensional Lanczos filter (Duchon 1979; Gulev et al. 2002). First, the flux products were interpolated on a half-degree grid and then the Lanczos filtering has been applied for 36 directions with the cutoff radius varying from 166 to 455 km.

Figure 2 shows an example of the methodology used and demonstrates the time-dependent influence of sampling on sea–air flux estimates in the Labrador Sea, playing a key role in driving the Atlantic general circulation (e.g., LabSea Group 1998). Figure 2a presents sensible heat fluxes in the Labrador Sea for January 1978 recomputed from the NCEP–NCAR reanalysis data and those computed from VOS data. The monthly mean sensible heat flux obtained from VOS is 6 W m^{-2} smaller than that obtained from NCEP–NCAR if only data collocated in time are used. However, the total monthly mean sensible heat flux from NCEP–NCAR, using all 124 six-hourly values available, is 160 W m^{-2} higher than that derived from all the 14 VOS reports. This highlights the significance that the effect of inadequate sampling may have. Note that we selected one of the best-sampled months (14 samples) during the period after 1973 when the typical sampling density in this region lies between zero to seven reports per month (Fig. 2b). During the period 1948–73 observations at OWS B (56.5°N , 51°W) contributed 80% to 90% of the total number of reports and nearly 98% of all winter observations in the Labrador Sea. Termination of the OWS B in 1973 resulted in a considerable

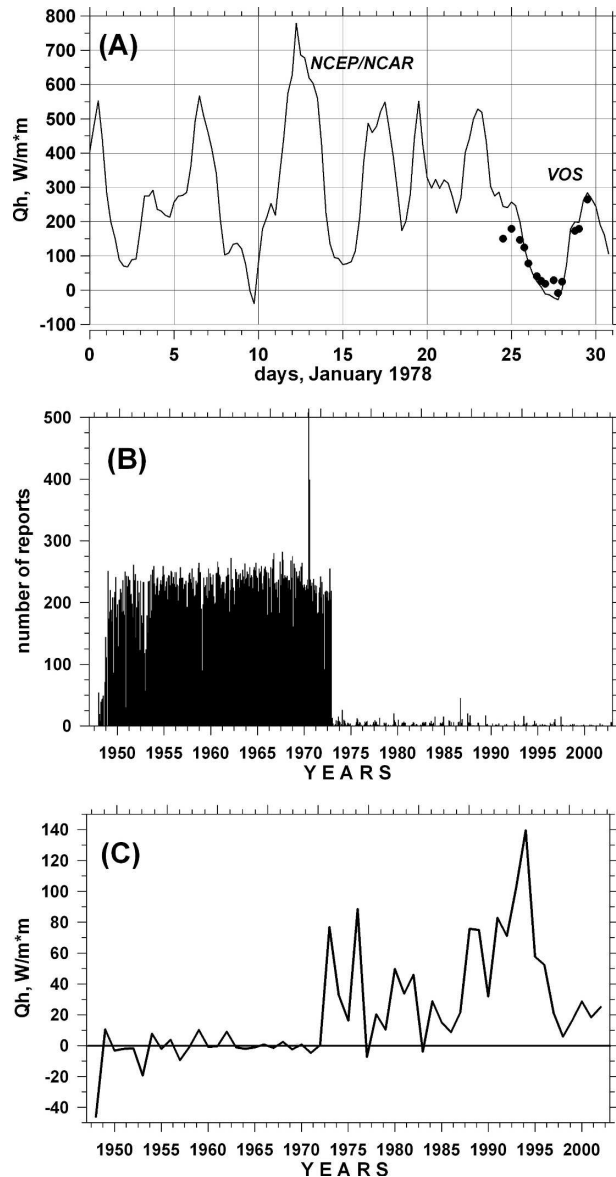


FIG. 2. (a) Time series of sensible heat flux derived from 6-hourly NCEP–NCAR data and available VOS reports for January 1978 in the 2° box in the Labrador Sea, (b) variability in the monthly number of reports in the Labrador Sea region restricted by the lines along 56°N and 45°W , and (c) winter (JFM) differences between the ICOADS-like sampled and original NCEP–NCAR net fluxes in the $4^\circ \times 4^\circ$ box around OWS B.

drop in the number of VOS reports, especially in winter (Fig. 2b). Figure 2c shows winter [January–March (JFM)] differences between the fully sampled NCEP–NCAR sensible heat fluxes and those derived from the NCEP–NCAR 6-hourly values matching the VOS observations in a 4×4 degree box around the OWS B. During the period 1948–72 differences vary within 10 W m^{-2} with an average of 2 W m^{-2} and a standard deviation

tion (std) of 8 W m^{-2} . However, after 1973, differences are positive for most years with a mean of 51 W m^{-2} and a std of 34 W m^{-2} . Figure 2c shows that even for long-term averages, sampling biases may amount to as much as 50 W m^{-2} in highly variable areas such as Labrador Sea. Similar conclusions can be drawn from the data of the LabSea experiment (LabSea Group 1998; Renfrew et al. 2002). During February–March 1997 the research vessel (RV) *Knorr* collected several hundred meteorological observations in the Labrador Sea (Renfrew et al. 2002). We selected for the same period only those VOS reports that do not belong to RV *Knorr* dataset (eight reports) and obtained an underestimation of the *Knorr* values (including additional eight VOS samples) of 105 and 71 W m^{-2} for sensible and latent heat, respectively.

d. Estimation of sampling uncertainties

Sampling uncertainties depend on the sampling density and on the method used to carry out the subsampling. If VOS-like sampling is simulated according to the real time of the VOS reports, the squared difference between the monthly means of the regularly sampled variables or fluxes (x_o) and VOS-like sampled variables (x_u) gives the squared *total sampling uncertainty*, χ^2 :

$$\chi^2 = (x_o - x_u)^2. \quad (1)$$

Alternatively, the simulation of sampling can be carried out by randomly choosing n NWP data, n being the number of VOS observations for one particular box and month. This method has been traditionally used to estimate sampling uncertainties (e.g., Kidson and Trenberth 1988; Gulev et al. 2003b; Chang 2005). This procedure yields the *random sampling error*, δ , which is only associated with the number of samples available; the actual UTC times of observations are not considered. Following Gulev et al. (2003b), for each month and $2^\circ \times 2^\circ$ box we repeated this procedure 20 times, yielding 20 estimates of δ_i , $i = 1, \dots, 20$. The value $\langle \delta_i^2 \rangle^{1/2}$, with $\langle \rangle$ being the averaging operator, gives an estimate of the monthly random sampling error. Sensitivity experiments for selected years show that dependence of δ on the number of simulations becomes 3–8 times weaker for $i = 7$ and becomes negligible for $i > 14$. For ECF, for which no atmospheric data were used to constrain the model, only random sampling errors were estimated. Although the model may quite successfully simulate basic characteristics of synoptic variability, it cannot be expected to reproduce particular synoptic events after some days when the impact of initial conditions goes to zero (loss of predictability).

The difference between the squared total sampling error and the squared random sampling error, that is,

$$\xi^2 = \chi^2 - \delta^2, \quad (2)$$

represents the squared *actual sampling error*, which is not random. It partly accounts for the fair-weather bias, which is associated with the fact that ships tend to avoid stormy conditions. However, for many poorly sampled regions actual sampling errors are dominated by the fact that ships tend to pass grid boxes during limited periods only (hence, e.g., providing several measurements of the same synoptic situation). In this case the actual VOS-like sampling effectively provides a smaller number of *independent* observations in comparison with the random sampling procedure.

Estimates of the total and random sampling errors are valid only for grid cells with at least one report per month. For the fully unsampled boxes, sampling uncertainty is represented by the spatial interpolation error, $\gamma(G)$, which depends on the interpolation operator G used for the particular flux climatology to provide users with gap-free fields. This error was quantified by comparing the monthly means of the regularly sampled NWP variables and those implied by the interpolation procedure for the grid cells with no observations in ICOADS. Note that skills of the interpolation procedure and, therefore, the associated uncertainty will also depend on the sampling errors in the neighboring locations.

For individual flux-related variables there are two kinds of sampling uncertainties. The first one corresponds to individual parameters such as air temperature (Fig. 1a). The second one arises since more than one parameter has to be available simultaneously in order to estimate surface fluxes (Fig. 1b). We will call these two estimates “parameter-sampling errors” (χ_p , ξ_p , δ_p , γ_p) and “flux-sampling errors” (χ_f , ξ_f , δ_f , γ_f). All sampling uncertainties described above were estimated for meteorological variables as well as for air–sea fluxes (Q_h , Q_e , SW, LW, net heat flux) using 6-hourly data from NCEP–NCAR, ERA-40, and ECF and ICOADS reports for the corresponding periods. Our focus will be on the NCEP–NCAR-based results, which we will compare with 10 yr of ERA-40 and ECF.

4. Climatological sampling errors in individual variables and fluxes from NCEP–NCAR reanalysis data

a. Random sampling errors

The random sampling errors in air temperature, $\delta_p(T_a)$, scalar wind speed, $\delta_p(V)$, and total cloud cover, $\delta_p(C_n)$ for January and July for the period 1948–2002 (Fig. 3) represent the joint effect of the sampling density and the intensity of synoptic-scale variability of in-

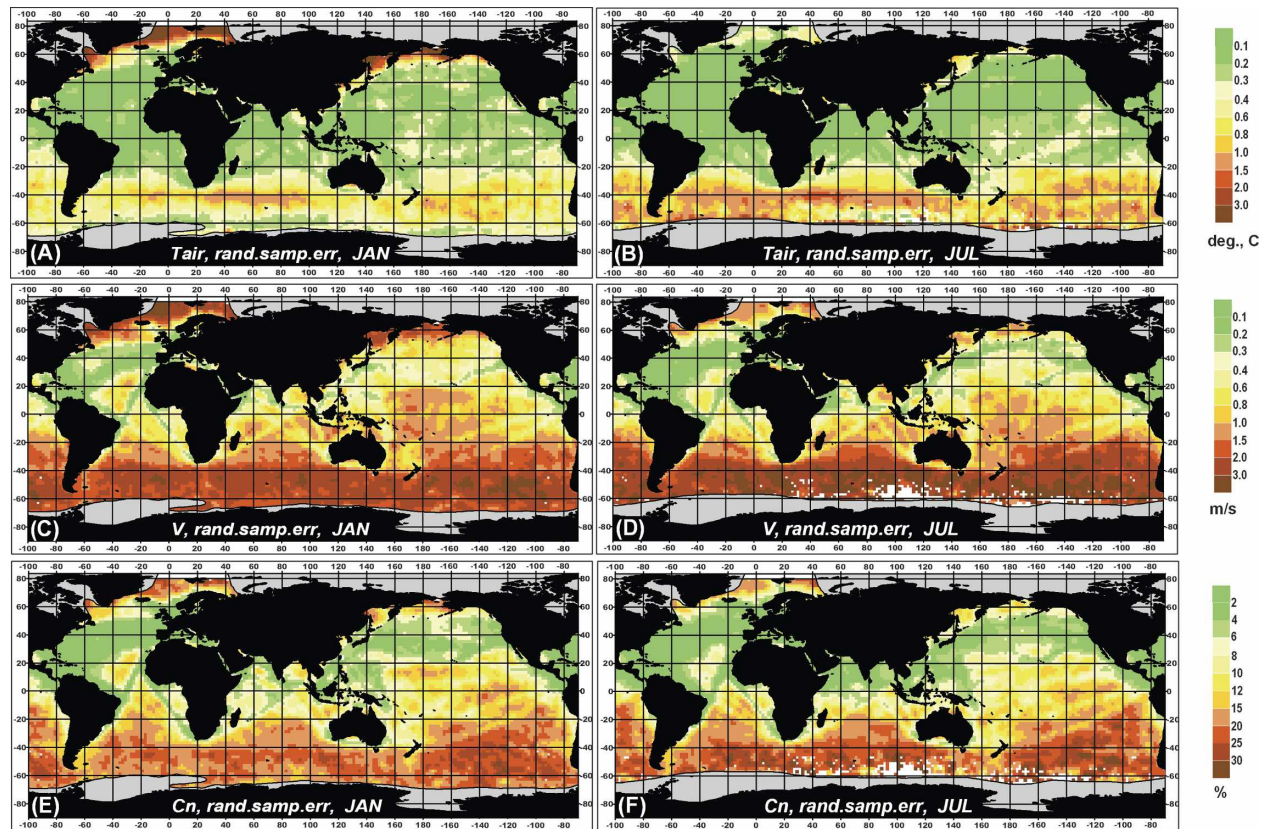


FIG. 3. Random sampling errors δ_p in (a), (b) air temperature, (c), (d) scalar wind speed, and (e), (f) total cloud cover for (a), (c), (e) January and (b), (d), (f) July.

dividual parameters. The largest winter values of $\delta_p(T_a)$ amount to 3°C in the poorly sampled Labrador and Greenland–Iceland–Norwegian (GIN) Seas in the Atlantic and Okhotsk and Bering Seas in the Pacific, as well as in the Southern Ocean. The smallest values of $\delta_p(T_a)$ ranging within 0.2°C are observed along the best-sampled ship routes in the midlatitudinal and tropical oceans, which are characterized by weaker synoptic variability. The largest random sampling errors in the scalar wind speed are higher than 3 m s^{-1} in the cold season in the North Atlantic subpolar latitudes and in the Southern Ocean and drop to $0.1\text{--}0.4\text{ m s}^{-1}$ in well-sampled areas. Random sampling errors in the total cloud cover vary from several percent in well-sampled regions to $15\%\text{--}25\%$ in the poorly sampled subpolar regions of the Northern Hemisphere and Southern Ocean. Random sampling errors in specific humidity (not shown) generally follow the spatial distributions of $\delta_p(T_a)$ and have maximum values of $2\text{--}3\text{ g kg}^{-1}$ in the poorly sampled regions of the subpolar North Atlantic in winter.

The ratio between the “flux sampling” errors (δ_f) and the “parameter sampling” random errors (δ_p) var-

ies from 1.5 to 2 in the Northern Hemisphere midlatitudes to values slightly higher than 1 in the Southern Ocean. Comparison of zonal averages of (δ_f) and (δ_p) (Fig. 4) shows that values of $\delta_f(T_a)$ are larger than those of $\delta_p(T_a)$ by $0.1^\circ\text{--}0.3^\circ\text{C}$ in the Northern Hemisphere midlatitudes and by $0.08^\circ\text{--}0.15^\circ\text{C}$ in the Southern Ocean. The highest values are typically found during the cold season. For the wind speed, the largest differences (about 0.8 m s^{-1} in winter and $0.3\text{--}0.5\text{ m s}^{-1}$ in summer) are observed in the Northern Hemisphere midlatitudes. Zonally averaged differences between the values of $\delta_f(C_n)$ and $\delta_p(C_n)$ (not shown) vary from 2% to 8% with the maximum found in the Northern Hemisphere midlatitudes. A qualitatively similar distribution is observed for specific humidity typically in the range $0.1\text{--}1.8\text{ g kg}^{-1}$ (not shown).

The largest random sampling errors in sensible heat flux (Figs. 5a,b) during boreal winter amount to as much as 60 W m^{-2} in the Labrador and GIN Seas in the North Atlantic as well as in the western marginal seas of the Pacific. For the latent heat flux (not shown) the largest sampling errors are observed in the Southern Ocean, varying from 40 to 60 W m^{-2} during the cold

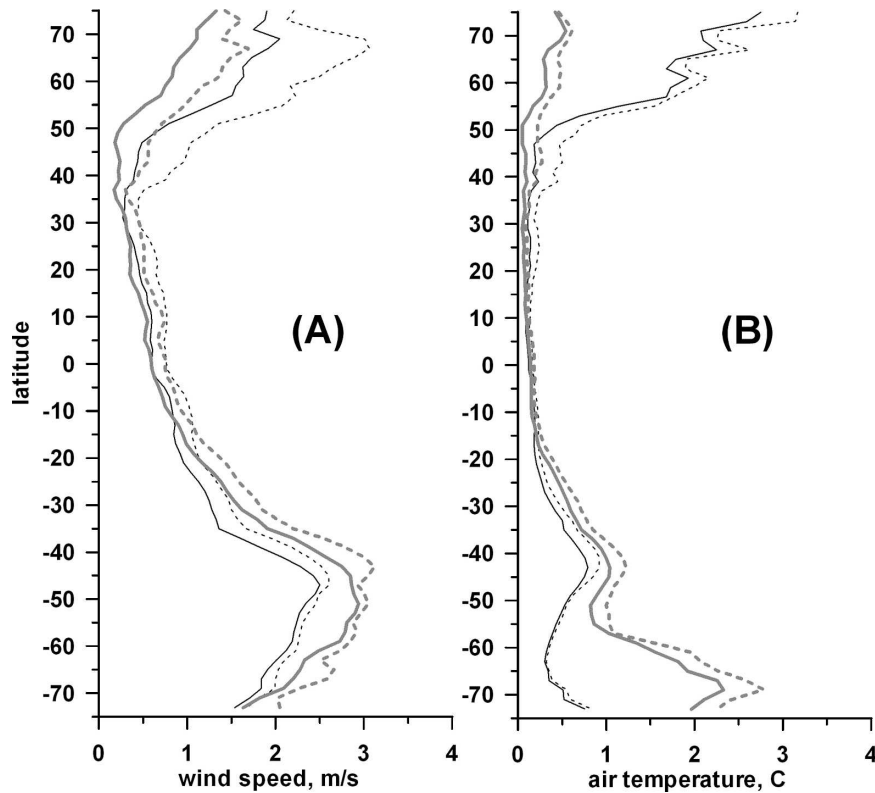


FIG. 4. Zonally averaged “parameter” (solid lines) and “flux” (dashed lines) sampling random errors in (a) scalar wind speed and (b) air temperature for January (thin black lines) and July (thick gray lines).

season and being 20%–30% smaller in austral summer. Random sampling errors in radiative fluxes (Figs. 5c–f) are smaller than those in the turbulent fluxes and amount to 10–15 W m^{-2} for SW radiation and to 20 W m^{-2} for LW flux. The highest $\delta_f(\text{SW})$ values ($>15 \text{ W m}^{-2}$) are observed in the Southern Ocean during austral summer and in the North Atlantic subpolar latitudes during boreal summer, when the actual values of SW radiation are relatively high. Locally, the high values of $\delta_f(\text{LW})$, ranging from 20 to 40 W m^{-2} , are observed in the subpolar latitudes of the North Atlantic during boreal winter. Random sampling errors in the net heat flux (not shown) were estimated from the individual values of the net fluxes. The largest values may amount to as much as 40–80 W m^{-2} and observed during the cold season in the subpolar North Atlantic and North Pacific and in the Southern Atlantic and Indian Oceans.

b. Total and actual sampling errors

The total sampling errors, χ , which take into account the actual VOS-like sampling, show spatial patterns very similar to those obtained for random sampling er-

rors (not shown); their magnitude, however, is generally larger. Zonal differences between the total and random sampling errors (Fig. 6) represent the actual sampling uncertainties ξ according to Eq. (2). The largest values of $\xi_p(T_a)$ amount to 2°–3°C in winter in the Northern Hemisphere subpolar latitudes and from March to October in the Southern Ocean south of 60°S. For the wind speed the largest differences of about 1.5 m s^{-1} are observed in the subpolar latitudes of the Northern Hemisphere during the cold season, reaching a maximum of 2.5–3 m s^{-1} in the subpolar northwest Atlantic. In the Southern Ocean $\xi_p(V)$ varies from 1 to 1.7 m s^{-1} and does not show as clear a seasonal dependence as was found in the Northern Hemisphere.

Figure 6c shows that the actual sampling adds up to 20 W m^{-2} to the zonal random sampling errors in Q_h in the Northern Hemisphere subpolar latitudes during boreal winter. January and July maps of the actual sampling errors in Q_e (Figs. 7a,b) are somewhat different from those for the random sampling errors. The highest winter values of $\xi_f(Q_e)$ of up to 30 W m^{-2} are observed in the subpolar North Atlantic, in the Gulf Stream area and in the midlatitudinal northwest Pacific, implying total sampling errors from 50 to 80 W m^{-2} . The highest

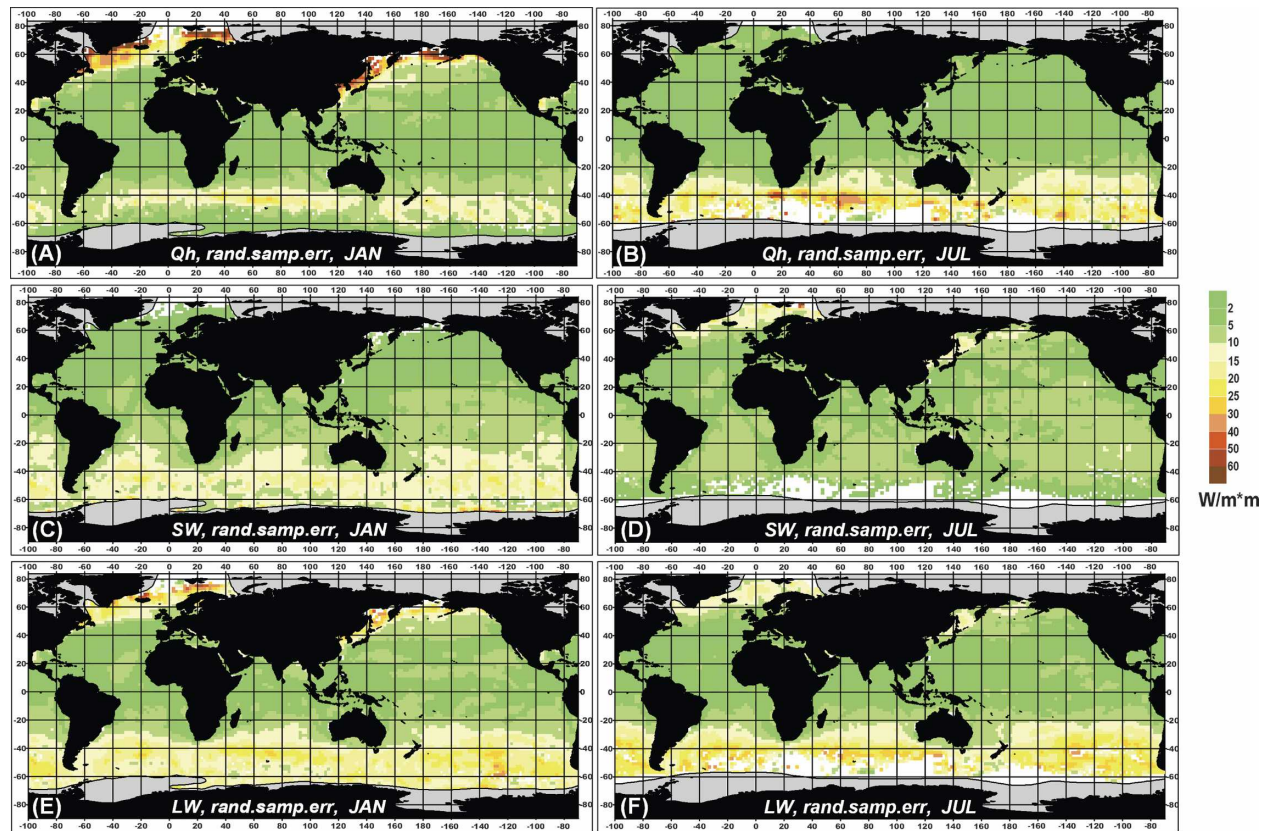


FIG. 5. Random sampling errors δ_f in (a), (b) sensible heat flux, (c), (d) SW radiation, and (e), (f) LW radiation for (a), (c), (e) January and (b), (d), (f) July.

July values ($30\text{--}50\text{ W m}^{-2}$) are evident in the South Pacific and Atlantic Oceans. In many poorly sampled boxes of the Southern Ocean (less than three–five reports per month) the values of $\chi_f(Q_e)$ and $\delta_f(Q_e)$ are quite close to each other. This shows that the total sampling error in this area is to a lesser degree determined by the actual sampling. The errors in Q_h show a similar behavior in this area (not shown). The highest winter values of $\xi_f(LW)$ (Figs. 7c,d) range from 15 to 30 W m^{-2} in the subpolar North Atlantic and from 10 to 20 W m^{-2} in the North Pacific. The highest July values of $\xi_f(LW)$ of up to 30 W m^{-2} are found in the Southern Ocean. Zonally averaged actual sampling errors in SW radiation (Fig. 6d) show maxima of 10–20 W m^{-2} during the respective summer seasons of both hemispheres. Figure 8 shows the total sampling errors in the net sea–air heat flux for January and July. Representing the combined effect of the uncertainties in all flux components, the values of $\chi_f(Q_n)$ reach a maximum of 70–100 W m^{-2} in the subpolar North Atlantic in boreal winter. Maximum winter values of $\chi_f(Q_n)$ in the North Pacific are smaller and range from 30 to 50 W m^{-2} . Total sampling errors in the net heat flux in the South-

ern Ocean vary from 40–60 W m^{-2} in January to 50–80 W m^{-2} in July. The smallest values of $\chi_f(Q_n)$ are observed along well-sampled ship routes and lie within 10–15 W m^{-2} .

c. Dependencies of the sampling uncertainties on the number of samples and the intensity of synoptic variability

Understanding how sampling errors depend on the number of samples may provide insights into how many measurements are needed to obtain heat fluxes of a prescribed accuracy. Weare and Strub (1981) and Weare (1989) for the Tropics and Cayan (1992) for the midlatitudes found that about 10 observations per month per 5° box guarantee unbiased estimates of turbulent fluxes. However, Legler (1991) accounted for the intramonthly variability and concluded that the accurate computation of, for example, monthly mean wind speed in midlatitudes requires from 90 to 170 reports. Chang (2005) analyzed the sampling impact on monthly sextile statistics of ICOADS SLP and found that 20–30 observations are needed for unbiased esti-

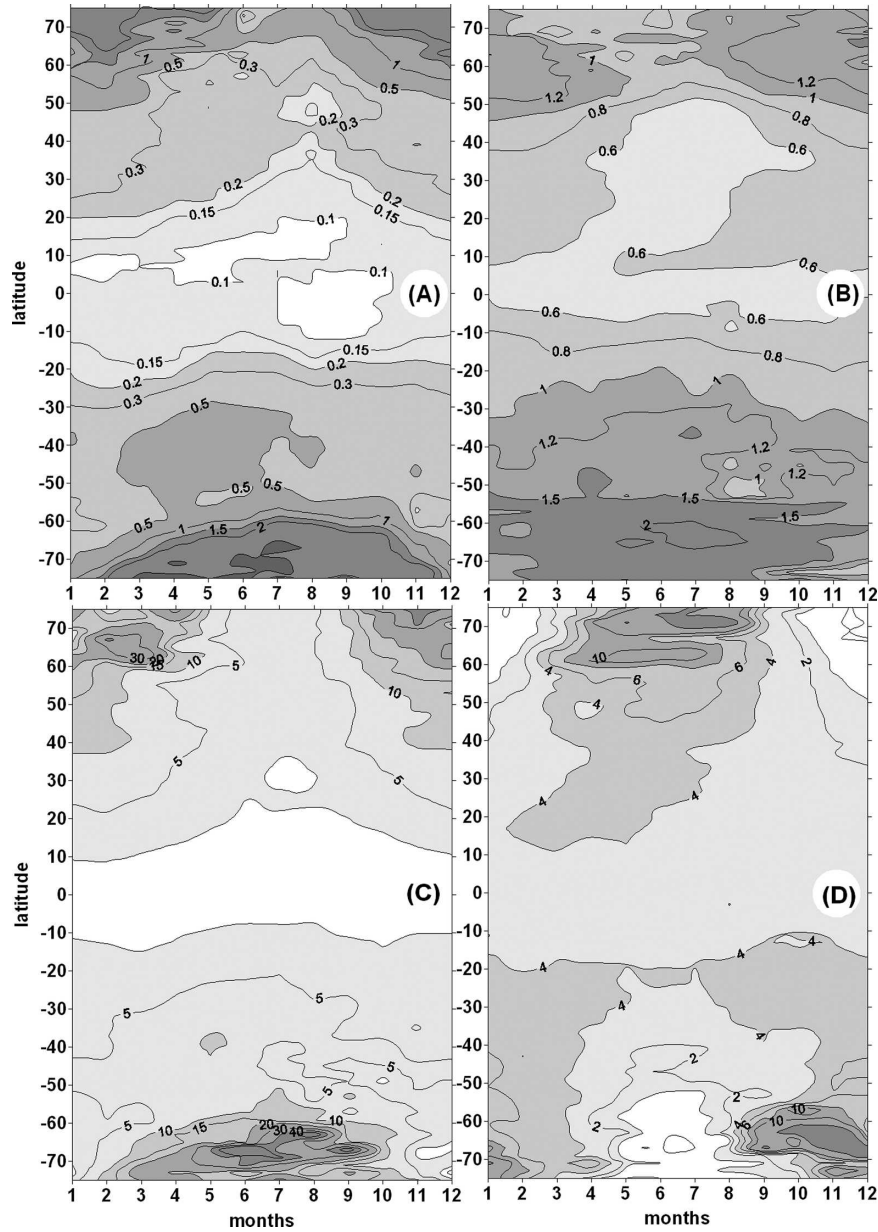


FIG. 6. Zonal seasonal differences between the total and random sampling errors, representing the actual sampling uncertainties in (a) air temperature ($^{\circ}\text{C}$), (b) scalar wind speed (m s^{-1}), (c) sensible heat flux (W m^{-2}), and (d) SW radiation (W m^{-2}).

mation. We sorted all monthly estimates of χ and δ for the 55-yr period into classes of the number of samples from 1 to 120. The number of estimates available reduces from $3\text{--}5 \times 10^5$ for $1 < n < 5$ down to 1000–5000 for the $n > 50$. The dependence of sampling errors on n shows quite a large scatter, implying large uncertainties of the approximation. To minimize these uncertainties, sampling errors for parameters and fluxes were normalized:

$$\hat{\chi}(x) = \chi(x)/\sigma(x), \quad \hat{\delta}(x) = \delta(x)/\sigma(x), \quad (3)$$

where $\sigma(x)$ is the intramonthly standard deviation of the variable x , estimated from the regularly sampled NWP data. Normalized sampling errors decrease with the number of samples from a value of about 1 for $n = 1$ (for both δ and χ) to 0.06–0.1 (for δ) and 0.15–0.20 (for χ) for $n > 40$ (Fig. 9). For large enough n the total sampling errors are typically 2–3 times higher than random errors. Dependencies of the normalized sampling errors on the number of samples n_x can be approximated as follows:

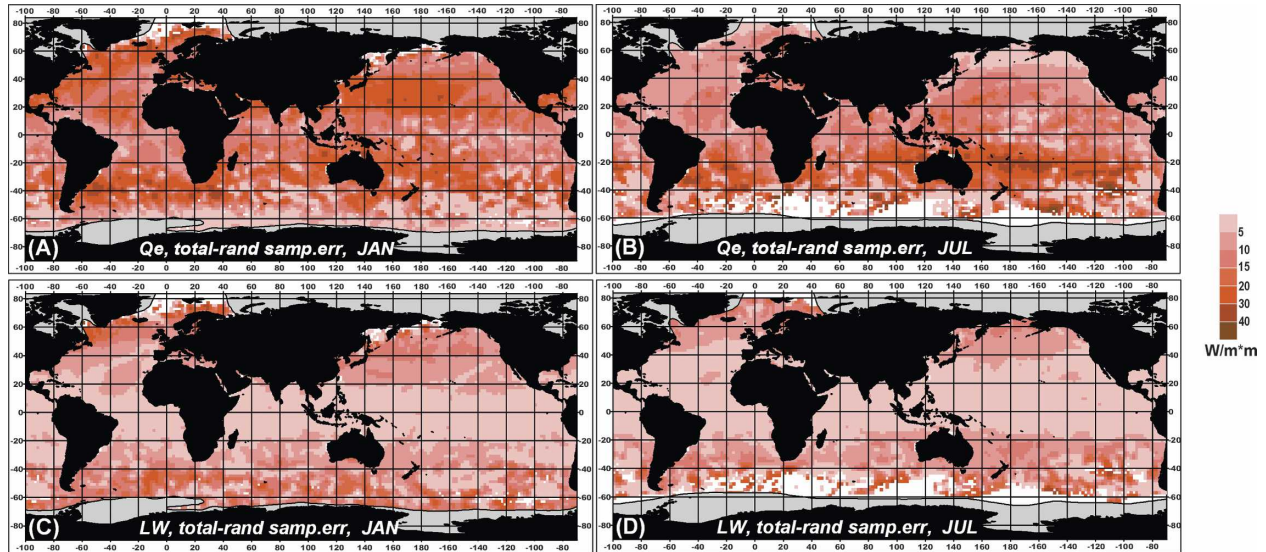


FIG. 7. Differences between the total and random sampling errors, representing the actual sampling uncertainties in the (a), (b) latent heat flux and (c), (d) LW radiation for (a), (c) January and (b), (d) July.

$$\ln \left\{ \begin{array}{c} \hat{\chi}(x) \\ \hat{\delta}(x) \end{array} \right\} = \left\{ \begin{array}{c} a_x \\ a_\delta \end{array} \right\} \ln n_x + \left\{ \begin{array}{c} b_x \\ b_\delta \end{array} \right\}, \quad (4)$$

where a_x , a_δ , b_x , and b_δ are empirical coefficients that were determined by a least squares approach using all data. Numerical values for these coefficients are given

in Table 1 for all sea–air flux components and can be further used for the estimation of the sampling errors for a given number of reports per month and 2° boxes. The fastest decrease of sampling errors with n is found for the SW radiation and the slowest for sensible heat flux, implying that less observations are needed for the

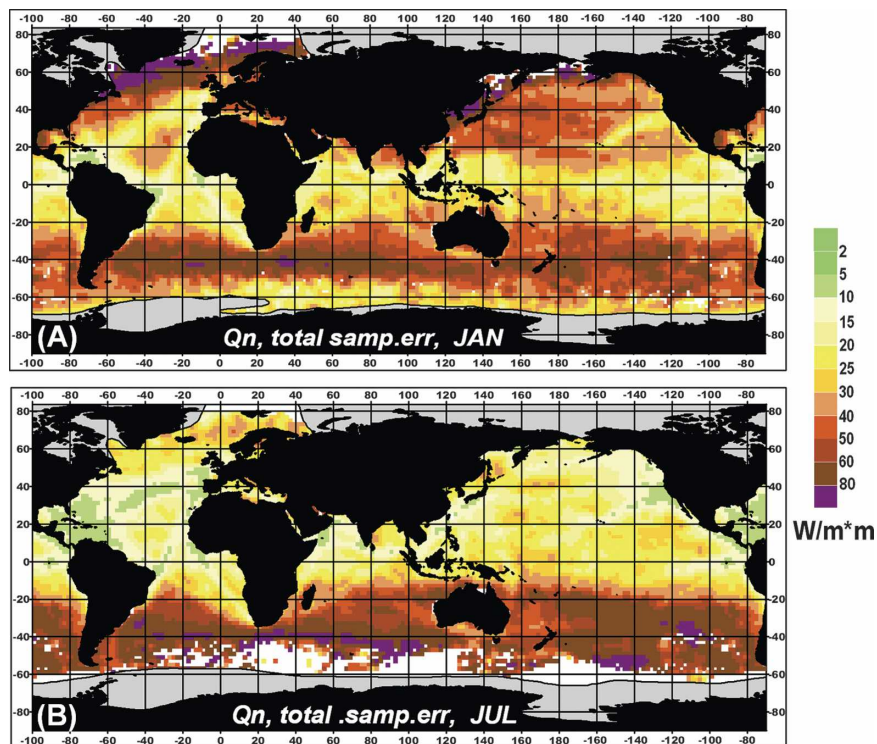


FIG. 8. Total sampling errors in the net heat flux for (a) January and (b) July.

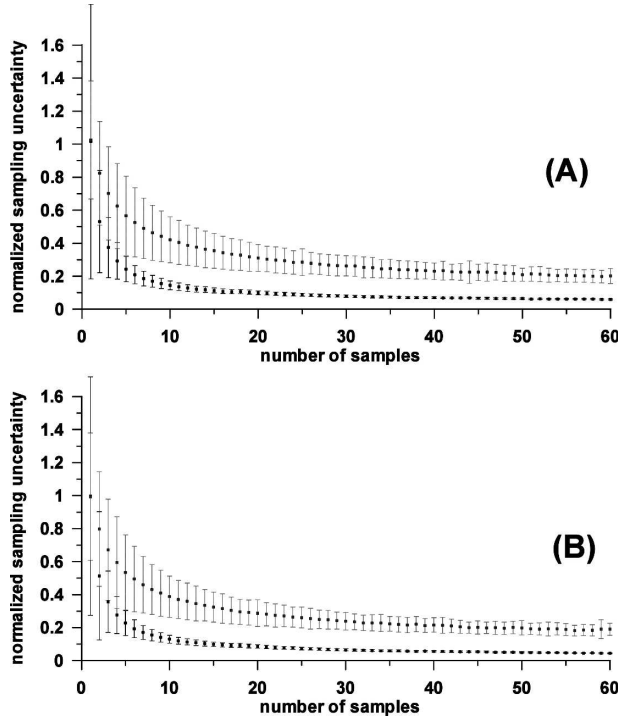


FIG. 9. Dependencies of the normalized total (gray) and random (black) sampling errors (3) on the number of samples for the (a) wind speed and (b) latent heat flux. Error bars show the standard deviations for individual bins.

adequate estimation of monthly SW radiation in comparison to more highly variable turbulent fluxes, which require more reports for obtaining the accurate monthly mean.

d. Uncertainties due to interpolation and extrapolation in data-void areas

Being also formally associated with sampling density, interpolation errors, γ , by nature are somewhat different from the sampling uncertainties described above. They depend on the procedure of spatial interpolation applied, on the arrangement of data-void locations, and the spatial correlation of variables being analyzed. Ad-

ditionally, sampling uncertainties in areas surrounding data-void grid cells have an impact on the interpolation errors (e.g., Kent et al. 2000). Figures 10a–d show the interpolation uncertainty for the wind speed and sensible heat flux, estimated by comparing regularly and VOS-like sampled NCEP–NCAR monthly fields. For wind speed, the largest January values of $\gamma_r(V)$ exceeding 2 m s^{-1} are observed in the Labrador and GIN seas in the Atlantic, whereas in the Sea of Okhotsk and subpolar northwest Pacific $\gamma_r(V)$ may amount to 3 m s^{-1} . During austral summer the interpolation error can amount to $2\text{--}3 \text{ m s}^{-1}$ in the southeast Pacific and southwest Indian Ocean. The interpolation errors in sensible heat flux (Figs. 10c,d) vary from several watts per meter squared in the Tropics and midlatitudes during the warm season to 60 W m^{-2} in the subpolar latitudes of the North Atlantic and North Pacific and to $30\text{--}50 \text{ W m}^{-2}$ in the Southern Ocean during the cold season. In the poorly sampled Labrador Sea and northwestern Pacific, interpolation errors can be 10%–30% higher than sampling errors (Figs. 5–8), largely explaining the total uncertainty associated with inadequate sampling.

Interpolation uncertainties in the latent heat flux (not shown) are 10% to 30% higher than those for the Q_h in the midlatitudes and Tropics and 20%–50% smaller in the subpolar latitudes. Zonally averaged interpolation uncertainties in the SW and LW radiation (Fig. 11a) vary within 5 W m^{-2} in the Tropics and midlatitudes and increase to about 20 W m^{-2} in the subpolar latitudes of the Northern and Southern Hemispheres during the cold season. Figure 11b shows that from 50°S to 40°N total sampling uncertainties in the net heat flux are 30%–60% larger than interpolation errors. However, in the northern and southern subpolar latitudes interpolation errors may become larger than the sampling errors with the locally highest differences in the poorly sampled and semi-isolated Labrador and GIN Seas and in the Sea of Okhotsk.

How sensitive are these results to spatial interpolation schemes? We tested successive correction, used by Josey et al. (1999) and da Silva et al. (1994) and krigging, applied by Lindau (2000), for the northwest and southern Atlantic. Table 2 compares the values of γ , obtained using these methods and the modified method of local procedures used in this study. Area-averaged estimates vary within the range of $\pm 10\%$ with better skills of the local procedure method in the Labrador Sea and of successive correction in the Southern Ocean. Krigging, although quite skillful on average in the Labrador Sea, may result in locally high uncertainties and therefore in higher standard deviations. Nevertheless, the spatial structure and the ratio between the interpo-

TABLE 1. Numerical values of the coefficients in approximation (4) for different sea–air flux components.

Air–sea flux component	a_δ	b_δ	a_x	b_x
Sensible heat	−0.597	−0.564	−0.408	−0.0138
Latent heat	−0.617	−0.541	−0.409	−0.0091
LW radiation	−0.616	−0.468	−0.418	−0.0043
SW radiation	−1.094	−1.310	−0.446	−0.0019
Net heat flux	−0.605	−0.550	−0.434	−0.0139

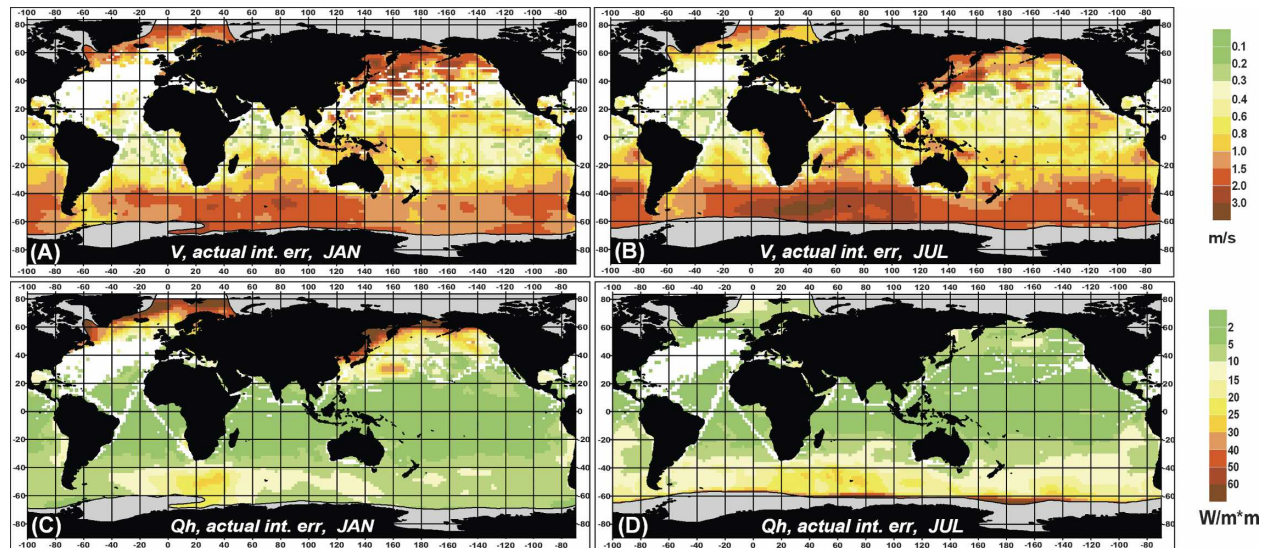


FIG. 10. Interpolation error in the (a), (b) scalar wind speed and (c), (d) sensible heat flux for (a), (c) January and (b), (d) July.

lation uncertainties and other sampling errors remain similar when estimated using different algorithms.

5. Comparison with alternative estimates of sampling uncertainties

a. Comparison with ERA-40 reanalysis

To compare estimates of sampling errors derived from the NCEP–NCAR reanalysis with alternative

data, we performed similar computations with ERA-40 reanalysis and compared the results for the period 1992–2001. This period in both reanalyses is largely affected by observational data, which were somewhat different for NCEP–NCAR and ERA-40 (Kalnay et al. 1996; Uppala et al. 2005). January and July maps of the total sampling uncertainty of ERA-40 sensible heat fluxes for 1992–2001 (Figs. 12a,b) are very similar to those for NCEP–NCAR for the same period. Because

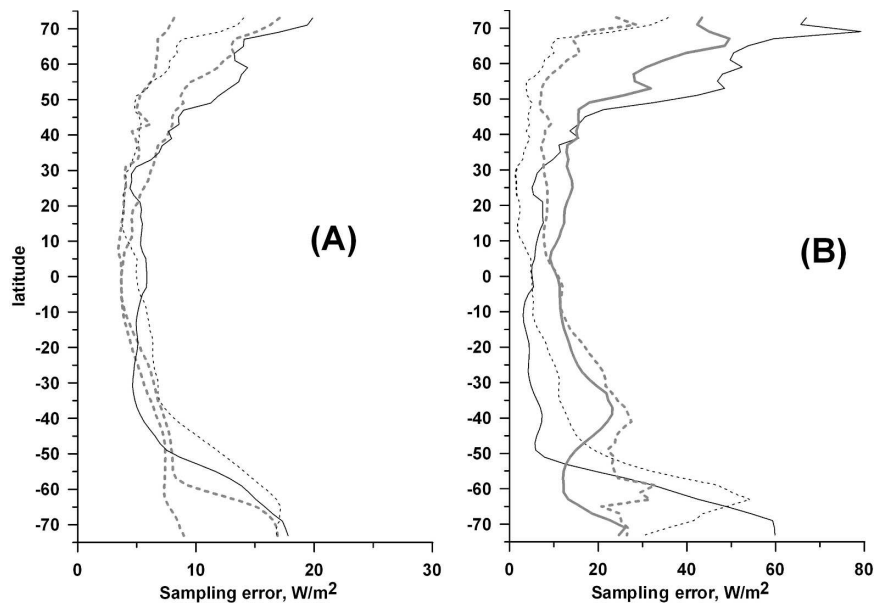


FIG. 11. (a) Zonally averaged interpolation errors in the SW radiation (thin black) and LW radiation (bold gray) for January (solid lines) and July (dashed lines). (b) Zonally averaged interpolation (thin black) and sampling (bold gray) errors in the net sea-air heat flux for January (solid lines) and July (dashed lines).

TABLE 2. Comparison of the area-averaged interpolation errors in sensible heat flux and LW radiation (W m^{-2}), obtained using different interpolation schemes in the northwest and southern Atlantic. First numbers are the interpolations errors, and second numbers are the area standard deviations.

Method of interpolation	Labrador Sea		Southern Atlantic	
	Q_h	LW	Q_h	LW
Local procedures	23.1/13.3	12.7/6.9	15.9/8.8	10.2/5.7
Successive correlation	25.3/17.6	14.1/6.8	13.7/7.4	9.4/5.2
Krigging	22.9/15.8	13.0/7.8	16.2/9.4	12.1/6.3

of poor sampling of the Southern Ocean in the 1990s, especially during austral winter, many grid cells in the southern Indian and Atlantic Oceans remained data void. In most areas the sampling uncertainty estimated from ERA-40 is somewhat larger than that derived from NCEP–NCAR (Figs. 12c,d). Global differences are 1.4 W m^{-2} for January and 1.6 W m^{-2} for July. Spatial distributions for the uncertainties in the latent flux derived from both reanalyses (not shown) are also very similar to each other with the higher values of

$\chi(Q_e)$ for ERA-40 (global differences 3.6 W m^{-2} for January and 4.1 W m^{-2} for July). Other flux-related variables and fluxes (SW and LW) also show slightly higher sampling uncertainties derived from ERA-40 data compared to NCEP–NCAR. This can most likely be explained by the higher-resolution model used for ERA-40 (T159 versus T62 in NCEP–NCAR) and, thus, potentially higher magnitudes of synoptic variability simulated in by the ERA-40 system. Our analysis shows that intramonthly standard deviations of surface variables and fluxes are systematically higher in ERA-40 in comparison to NCEP–NCAR by 1%–2% to 10%–20%. If we normalize the sampling uncertainties with standard deviations [Eq. (3)], the differences between the errors derived from two reanalyses become very small and do not show any regular pattern with the global mean deviations closely matching zero (Figs. 12e,f) and the pattern correlation exceeding 0.9, which is in line with the above explanation. Remarkably, the dependencies of $\hat{\chi}(n)$ derived from ERA-40 are hardly distinguishable from those obtained from NCEP–NCAR. In Fig. 13 we show approximations of $\hat{\chi}(n)$ derived ac-

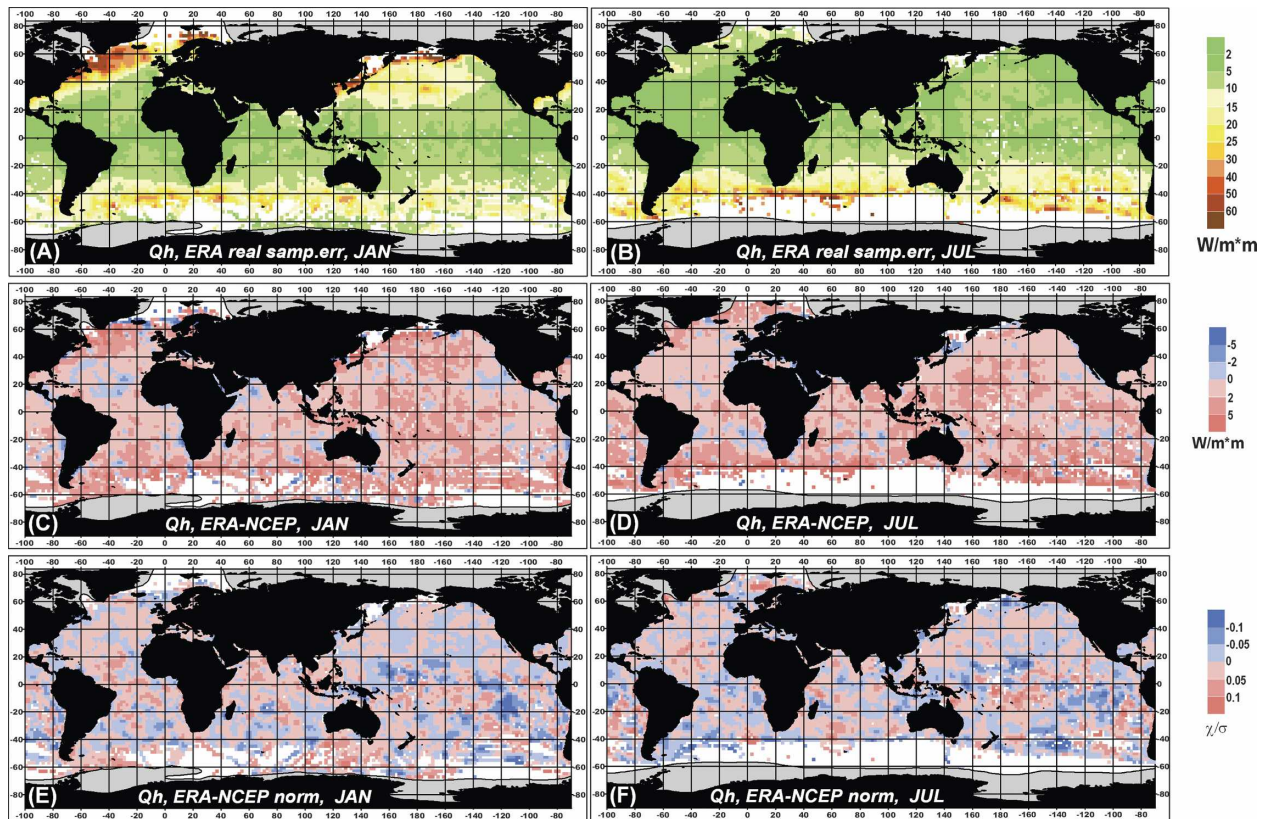


FIG. 12. Total sampling errors in sensible heat flux derived from the ERA-40 reanalysis for the period 1992–2001 for (a) January and (b) July, difference between the absolute total sampling errors in ERA-40 and NCEP–NCAR reanalyses for (c) January and (d) July, and the difference between the normalized total sampling errors in ERA-40 and NCEP–NCAR reanalyses for (e) January and (f) July.

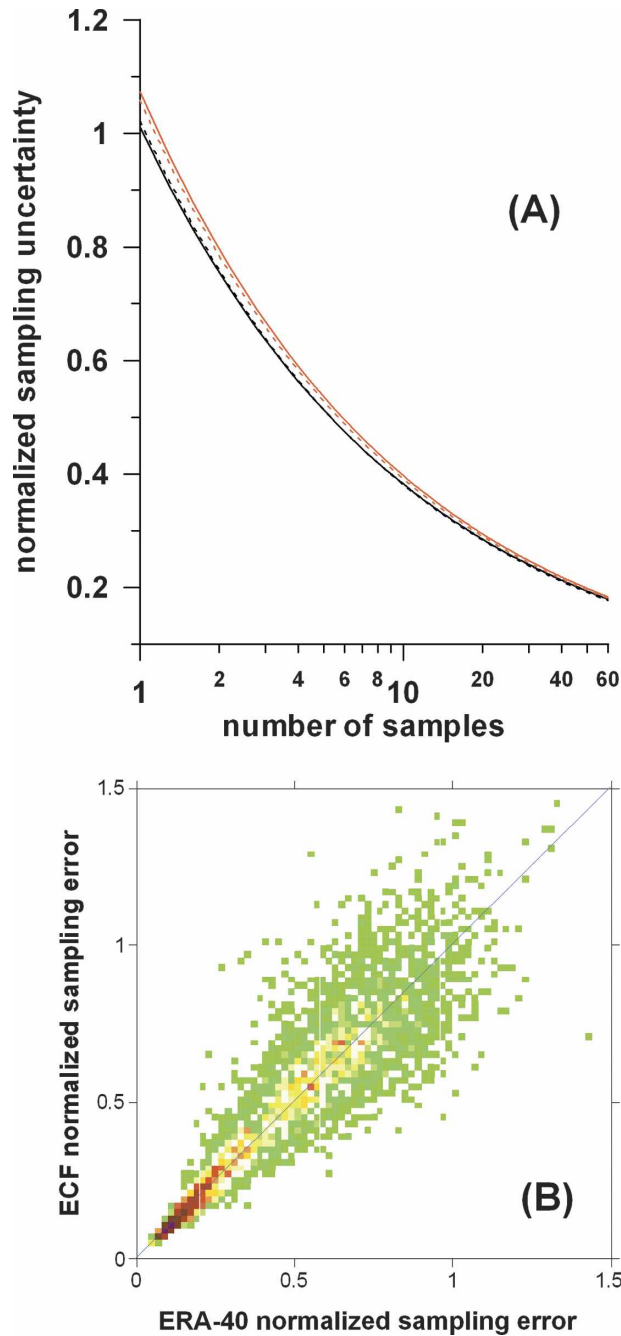


FIG. 13. (a) Dependencies of the normalized total sampling errors in sensible heat flux (red) and latent heat flux (black) derived from NCEP-NCAR (solid lines) and ERA-40 (dashed lines) data for the period 1992–2001 and (b) scatterplot of the normalized random sampling uncertainties in the latent heat flux derived from the ERA-40 and ECF data.

cording to Eq. (4) from NCEP-NCAR and ERA-40 for the period 1992–2001. They practically coincide with each other, implying that Eq. (4) is also valid for sampling uncertainties derived from ERA-40 data.

b. Comparison with ECMWF seasonal forecasts

Next, the impact of the use of actual data on the estimated sampling errors is investigated. VOS-like random subsampling is applied to wintertime seasonal integrations with the ECMWF model, ECF, in which no actual data were assimilated and the model used was very similar to the one used to carry out the ERA-40 reanalysis. As was mentioned above, the focus is on random sampling errors only, which are conceptionally better to understand for these seasonal integrations due to the fact that individual synoptic systems cannot be reproduced. Figures 14a,b show January maps of the random sampling uncertainties in the sensible and latent heat fluxes derived from ECF for the period 1992–2001. Again, the spatial patterns are quite comparable with those obtained for NCEP-NCAR and ERA-40 reanalyses with the highest uncertainties in poorly sampled and highly variable subpolar latitudes of the Northern Hemisphere and Southern Ocean. Differences between ECF and ERA-40 sampling errors (Figs. 14c,d) may locally amount to $10\text{--}15\text{ W m}^{-2}$, being larger than the differences between the estimates derived from the two reanalyses. For both sensible and latent heat fluxes systematically higher sampling uncertainties in ECF were identified in the midlatitudinal and subpolar North Pacific and systematically higher uncertainties in ERA-40 were found in the subpolar North Atlantic. This can be attributed to systematic errors of the ECMWF model climate in simulating synoptic variability for the two ocean basins (Jung 2005). This explanation is supported by the fact that the normalized sampling uncertainties are quite close to each other with a pattern correlation of 0.84 and with no regular pattern for the difference fields. Figure 13b shows the scatterplot of the normalized random sampling uncertainties in the latent heat flux derived from the ERA-40 and ECF data. Despite some reasonable scatter, the slope of the orthogonal regression closely matches 1 with intercept of nearly zero. Thus, analysis of two alternative NWP datasets supports our conclusions based on NCEP-NCAR data.

c. Comparison with independent VOS observations

To further assess how skillful the reanalyses are in simulating the magnitude of synoptic and mesoscale variability, we performed another regional comparison. To this end sampling uncertainties were estimated using exclusively those VOS data that were not included into ICOADS. Using the data collected under the SECTIONS program in the Newfoundland basin (Gulev 1994, 1999) we selected those boxes in which the SEC-

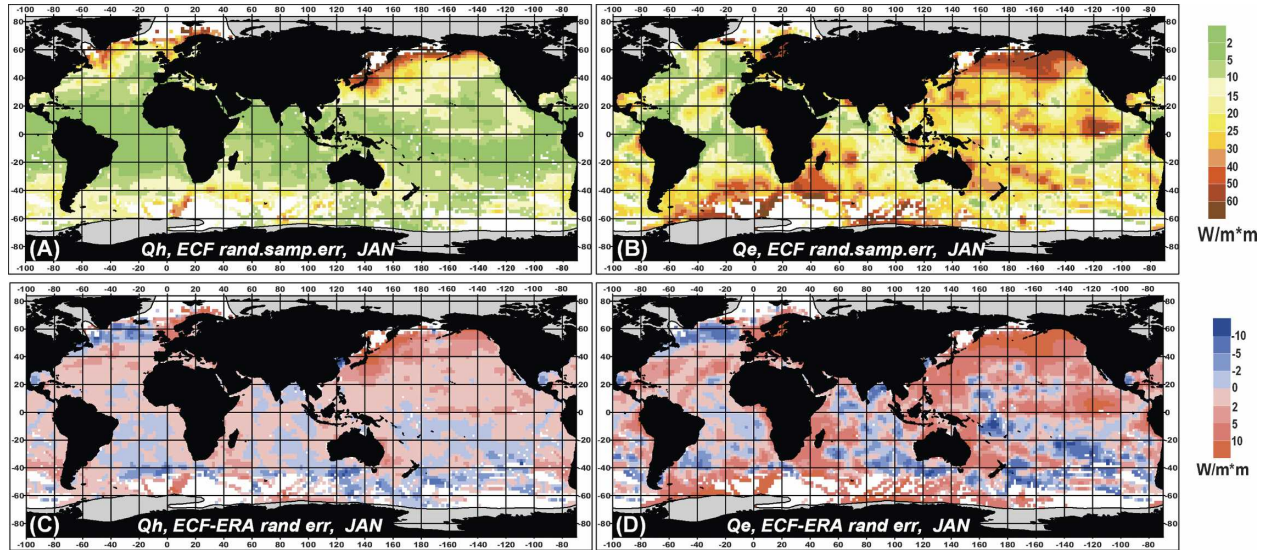


FIG. 14. Random sampling errors in (a) sensible and (b) latent heat flux derived from ECF for the period 1992–2001 for January and “ECF minus ERA-40” difference between the absolute random sampling errors in (c) sensible and (d) latent heat flux for January.

TIONS data provide more than 120 reports per month and the ICOADS provides less than 30 reports per month. Altogether six monthly time series were selected for 1987, 1988, and 1990 (Table 3). For these time series we performed the procedure of subsampling as described in section 3 by matching ICOADS reports with independent meteorological data from the SECTIONS collection. The flux estimates from SECTIONS data were computed using the same algorithms listed in section 3. Diagnosis of this small subset shows that the NCEP–NCAR reanalysis underestimates the magnitude of intramonthly variability by about 10% for wind speed and 3%–5% for other surface variables. On average, sampling errors obtained from SECTIONS data are $16\% \pm 5\%$ higher than those obtained from NCEP–NCAR reanalysis data (Table 3). If the effect of the magnitude of synoptic variability is removed by normalization, then the estimated sampling errors are very similar for the SECTIONS and NCEP–NCAR re-

analysis data (Table 3). Another independent dataset collected by the RV *Knorr* in Labrador Sea during one month of 1996 (LabSea Group 1998; Renfrew et al. 2002) also shows 6%–10% higher magnitude of synoptic variability in ship observations compared to NCEP–NCAR reanalysis data and, hence, implies higher actual sampling uncertainties. As with the other datasets described in this study, differences in estimated sampling uncertainties between the RV *Knorr* and NCEP–NCAR reanalysis data become small when normalized data are used.

6. Summary and discussion

We analyzed the impact of observed sampling uncertainties on the monthly mean fluxes (and flux-related variables) using 6-hourly NCEP–NCAR reanalysis data along with state-of-the-art bulk parameterizations. First, random sampling errors were obtained by ran-

TABLE 3. Estimates of the total sampling errors (W m^{-2}) and normalized sampling errors from NCEP–NCAR reanalysis and SECTIONS data for six monthly time series in Newfoundland basin.

Month, year	Number of reports		Total sampling error in sensible heat flux		Normalized sampling error in sensible heat flux	
	ICOADS	SECTIONS	NCEP	SECTIONS	NCEP	SECTIONS
Dec 1987	14	231	24.8	28.5	0.343	0.359
Mar 1988	26	496	11.7	13.4	0.213	0.218
Mar 1988	17	509	16.7	24.3	0.164	0.161
Jan 1990	11	188	26.9	29.7	0.353	0.339
Mar 1990	23	316	11.7	12.8	0.146	0.149
Apr 1990	18	454	8.4	13.1	0.162	0.168

TABLE 4. Averaged estimates of different sampling uncertainties in the net heat flux (W m^{-2}) for different World Ocean regions for January and July.

Regions	Random sampling uncertainty		Total sampling uncertainty		Interpolation uncertainty	
	Jan	Jul	Jan	Jul	Jan	Jul
Subpolar North Atlantic, 50°–70°N	23.7	6.9	34.8	12.2	47.6	11.3
Subpolar North Pacific, 40°–70°N	22.9	6.1	31.2	10.6	42.9	9.7
Midlatitudinal and subtropical North Atlantic, 30°–50°N	7.2	3.8	13.2	6.8	10.8	3.9
Midlatitudinal and subtropical North Pacific, 20°–40°N	6.9	3.6	15.3	8.1	12.1	4.3
Tropical oceans, 20°S–20°N	4.3	4.6	11.2	10.3	5.2	4.8
Southern Ocean, south of 30°S	16.7	22.3	22.4	28.5	23.2	27.9

domly subsampling the 6-hourly NCEP–NCAR data that had monthly frequency matching that found for VOS data. Second, total random errors were estimated by taking into account the actual dates and times at which the measurements were taken. By investigating (on a global scale) both random and total sampling models and by taking into account spatial interpolation errors as well, we obtained more advanced error estimates than earlier studies (Weare 1989; Cayan 1992), which used random sampling and VOS data only. Random sampling uncertainties in the flux-related variables amount to 2.5°–3°C for air temperature, 3 m s⁻¹ for the wind speed, 2–2.5 g kg⁻¹ for specific humidity, and 15%–20% for the total cloud cover in poorly sampled subpolar latitudes of the Northern Hemisphere and in the Southern Ocean. Total sampling errors are typically 20% to 30% higher than random ones, except for some poorly sampled areas, where the difference between them may be up to 60%. This difference can simply be explained by the fact that, on average, the number of independent observations during one particular month is increased if the available observations are randomly distributed throughout the month (see, e.g., Fig. 2). The largest random and actual sampling errors in surface sensible and latent heat fluxes in poorly sampled areas can amount to several tens of watts per meter squared with the total sampling uncertainty being from 20 to 50 W m⁻² higher than the random sampling error. The highest uncertainties of the interpolation into data-void grid cells are found in the subpolar latitudes of both hemispheres for the turbulent fluxes, where they are comparable with the sampling errors.

Table 4 summarizes estimates of different sampling uncertainties in the net heat flux for different ocean basins. It shows that actual sampling acts to increase sampling errors and that in high latitudes interpolation errors may be even higher than sampling uncertainties. The use of alternative reanalysis (ERA-40) and ECMWF seasonal forecasts changes the magnitude of sampling uncertainties with the spatial structure be-

ing comparable among different datasets. Normalization of the sampling errors with intramonthly standard deviations of variables and fluxes provides the possibility for their parameterization, which can be used for quantitative estimates of sampling uncertainties for a given number of samples in different regions.

Our estimates of sampling uncertainties can be compared with estimates of other sources of errors in surface heat flux fields and flux-related variables. Kent et al. (1999) and Kent and Berry (2005) report random observational errors for individual observations of 2–3 m s⁻¹ for wind speed and 0.7°–2°C for air temperature. To be comparable with sampling errors these values should be transformed into random observational uncertainties of monthly means, dividing them by $n^{1/2}$, with n being the number of observations, according to the central limit theorem for normally distributed random errors (Taylor 1982). For 2° × 2° boxes observational errors in monthly mean wind speed vary from 0.1 m s⁻¹ in well-sampled regions to 1–1.5 m s⁻¹ in poorly sampled areas. Thus, for $n < 10$ (most of subpolar North Atlantic and North Pacific and nearly the entire Southern Ocean) the random sampling error is higher than the random observational error. This is in agreement with Kent and Berry (2005) who compared their observational errors with the total (measurement plus sampling) errors of Brohan et al. (2003). For the Oberhuber (1988) climatology Gleckler and Weare (1997) found the largest uncertainties due to inaccuracy of parameterizations and observational errors of 30–40 W m⁻² in the net surface heat flux and from several W m⁻² to several tens W m⁻² in individual flux components. Our estimates of sampling errors in poorly sampled areas are of the same order of magnitude. Josey et al. (1999) reported the uncertainties associated with variable corrections to be from several W m⁻² to more than 15 W m⁻². In data-sparse areas our estimates of sampling uncertainties are several times higher, being smaller in well-sampled regions. Thus, our study reveals that sampling uncertainties in fluxes and flux-

related variables are of the same order of magnitude as the other uncertainties in air–sea fluxes in well-sampled regions and can be considerably higher than those due to other sources of uncertainty in poorly sampled areas. Our estimates can be used along with the estimates of observational errors to establish more reliable uncertainty ranges for climatological air–sea fluxes.

Not surprisingly it turns out that sampling uncertainties are largest in area with poor sampling such as the Labrador Sea. One might argue that due to the poor sampling what we see are simply uncertainties of the model used to carry out the NCEP–NCAR reanalysis (our “truth”). Although reanalysis products assimilate VOS reports and, thus, dependent on VOS data we nevertheless feel confident that what has been reported are good estimates of the sampling errors associated with VOS data. First, it should be pointed out that “poorly sampled” in terms of VOS data does not necessarily mean poor sampling in other observational data as well. A variety of different datasets are assimilated in reanalyses including radiosondes, dropsondes, and land-based synoptic observations. In fact, if these non-VOS data are considered as well then the Labrador Sea is not sampled that poorly after all. It is also worth pointing out the recent finding that the terrestrial observing system, of which VOS observations are a part, is not sufficient to skillfully reconstruct the observed Northern Hemisphere atmospheric circulation (Bengtsson et al. 2004). In this sense synoptic-scale systems in reanalysis data do not necessarily owe their existence to the assimilation of VOS data. Second, our study shows that it is possible to estimate sampling uncertainties using long model integrations alone (ECF) without assimilating any data. In fact, very similar estimates were obtained for ERA-40 and ECF. The results can be explained as follows: Suppose that no observational data were available at all (ECF). In this case the resulting “analysis” is solely based on a forecast and, hence, is strongly model dependent. Even though the “analysis” will have no skill in simulating actual weather events, our methodology still yields reasonable results. This can be explained by the fact that state-of-the-art models are very skillful in simulating the statistics of weather events, which is sufficient to get meaningful estimates of sampling errors (as illustrated by this study). The robustness of our estimates is further substantiated by computations performed with independent high-resolution data. This can be explained by the fact that state-of-the-art models are very skillful at simulating the statistics of weather events (Jung 2005). These data are characterized by higher absolute values of sampling uncertainties. However, the normalized

sampling errors derived from these data are closely correlated to those obtained from reanalyses. Moreover, reanalyses definitely underestimate synoptic variability of SST (especially before 1980s, before OI products became available) (Reynolds and Smith 1994) and, thus, short-term variability of fluxes, implied by SST fluctuations (Zolina and Gulev 2003). Thus, it is likely that our estimates of sampling errors are quantitatively underestimated by 10%–20%, but qualitatively show reliable patterns.

Similarly, the use of alternative bulk parameterizations may result in quantitatively different estimates of sampling uncertainties. Analyses of different schemes for surface turbulent and radiative fluxes (Blanc 1985; Dobson and Smith 1988; Bignami et al. 1995; Brunke et al. 2003; Eymard et al. 1999) as well as comparisons of different flux climatologies (Kent and Taylor 1995) show that despite quantitative differences, the spatial distribution and temporal variability are very close for the fluxes derived with different schemes. Thus, the relationships between sampling errors in different areas and different periods are likely to be quite similar for different algorithms. We recomputed our estimates with the turbulent scheme of Large and Pond (1982) and the LW radiation scheme of Josey et al. (2003). The differences in the absolute values of sampling errors were ranging within 10 W m^{-2} at the largest; that is, differences due to the use of different parameterizations are small compared to the sampling uncertainties. The normalized sampling errors as well as the coefficients of approximations (4) were very close to each other.

Of a special interest is to estimate the extent to which the difference between the total and random sampling errors can be attributed to the fair-weather bias. Our analysis and results of Kent and Taylor (1995) do not show significant changes in the major routes during severe weather. Typically, ships tend to slow down, though remaining on the same route and thus continue to report on GTS. The SHIPMET questionnaire (Gulev et al. 2003a) showed that stormy weather can influence the quality of the reports since the officers and mates do not precisely follow the observational guidelines. However, there was no indication of the systematic nature of biases. Fair-weather bias (if existent) should result in systematically too low wind speeds, but its impact on sea–air temperature and humidity is likely to be more random in nature. Thus, it can imply some underestimation of the turbulent fluxes. To test this hypothesis we analyzed actual values of the long-term differences between the fluxes computed from the regularly sampled and VOS-like sampled variables.

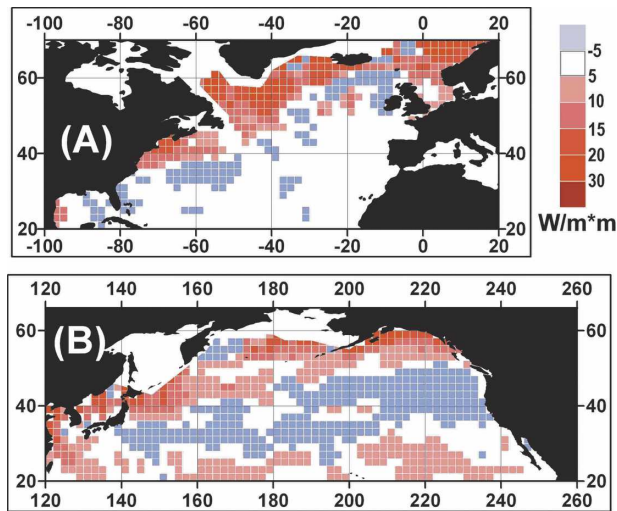


FIG. 15. Actual differences between sensible heat flux estimates computed from the regularly sampled variables and VOS-like sampled variables in the (a) North Atlantic and (b) North Pacific in winter. Values are shown only for the differences larger than interannual std dev of the differences.

Systematic effects with the magnitude larger than the standard deviation of the differences were found only in the northwest Atlantic and northwest Pacific for both sensible and latent heat fluxes (Fig. 15). Sensible heat flux is likely to be underestimated due to fair-weather bias by $15\text{--}30\text{ W m}^{-2}$ in the Labrador Sea and by $10\text{--}25\text{ W m}^{-2}$ in the northwest Pacific. Negative differences along the major midlatitudinal ship routes may reflect the oversampling of VOS data in comparison to 6-hourly NCEP–NCAR data. However, no regular pattern of either sign has been found in the Southern Ocean where the poor sampling is implied by seldom occurrence of the ship routes and not by rerouting due to severe weather.

Alternatively, the differences between the total and the random sampling errors may be associated with the higher number of *independent* observations in the random sampling model in comparison to the real VOS-like sampling, implying that the random sampling model is not fully adequate for the estimation of sampling errors. A detailed analysis of UTC report times and ship call signs shows that observations typically occur in series of several adjacent reports (see, e.g., Fig. 2). For the period 1960–79 we performed an alternative subsampling by randomly simulating adjacent series of observations according to the real sampling structure. The latter implies, for example, for the Labrador Sea, that five reports per month in 2° box in 23% of cases occur as one series of five adjacent reports, in 29%—as two series ($3 + 2$); in 16%—as two series ($4 + 1$); in

19%—as three series ($1 + 2 + 2$ or $1 + 1 + 3$); and only in 13%—as five single reports. This experiment showed about 20%–50% smaller difference between the total and random sampling error. Thus, a considerable part of the difference between the total and random sampling errors can be attributed to lower values of the number of *independent* samples in real VOS-like sampling.

Our analysis of the sampling impact on air–sea fluxes provides the background for the further evaluation of both NWP and VOS fluxes and may be useful for the development of new algorithms for the reconstruction of VOS-based flux time series. Methodologies based on EOFs and Kalman filtering developed for SST and SLP fields (Reynolds and Smith 1994; Smith and Reynolds 2003, 2004; Kaplan et al. 1998, 2000) provide more reliable time series than those derived through the raw averaging (Sterl 2001). However, being skilful in the reconstruction of anomalies, these methods may still be not effective enough in accounting for the uncertainties in climate means. Recently, Yu et al. (2004a,b) developed 10-yr daily fields of turbulent heat fluxes in the Atlantic Ocean using a variational objective analysis approach involving satellite and NWP data, and providing more accurate means and variability patterns. Further improvement of similar methodologies along with blending of different satellite products (Zhang et al. 2006) aimed at the development of long-term accurate air–sea flux time series will require estimates of the impact of sampling uncertainties on the climate variability of surface fluxes. This issue will be addressed in Part II (Gulev et al. 2007).

Acknowledgments. A considerable part of this study was done during the periodical work of SKG at IFM-GEOMAR during the last years. He thanks IFM-GEOMAR for its hospitality. We appreciate the help of our colleagues at IFM-GEOMAR: Karl Bumke, Ute Hecht, Thomas Martin, Anne Meyer, and Frauke Nevoigt. Special thanks to Scott Woodruff of NOAA/CDC (Boulder) and Steve Worley of DSS/NCAR (Boulder), who made available the ICOADS data for us. Discussions on different aspects of this work with Simon Josey and Liz Kent of SOC (Southampton), Andreas Sterl of KNMI (De Bilt), and Glenn White of NCEP (Camp Spring) were very helpful. Suggestions and criticism of anonymous reviewers and of the editor largely helped to improve the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft Sonderforschungsbereich SFB-460, Russian Ministry of Science and Education, under the World Ocean Federal Program and by Russian Foundation for Basic Research (Grant 05-05-64882).

REFERENCES

- Akima, H., 1970: A new method of interpolation and smooth curve fitting based on local procedures. *J. Assoc. Comput. Mach.*, **17**, 589–602.
- Bauer, M., A. D. Del Genio, and J. R. Lanzante, 2002: Observed and simulated temperature–humidity relationships: Sensitivity to sampling and analysis. *J. Climate*, **15**, 203–215.
- Bengtsson, L., K. I. Hodges, and S. Hagemann, 2004: Sensitivity of the ERA-40 reanalysis to the observing system: Determination of the global atmospheric circulation from reduced observations. *Tellus*, **56A**, 456–471.
- Bentamy, A., K. B. Katsaros, A. M. Mestas-Nuñez, W. M. Drennan, E. B. Forde, and H. Roquet, 2003: Satellite estimates of wind speed and latent heat flux over the global oceans. *J. Climate*, **16**, 637–656.
- Biglami, F., S. Marullo, R. Santorelly, and M. E. Schiano, 1995: Long-wave radiation budget in the Mediterranean Sea. *J. Geophys. Res.*, **100**, 2501–2514.
- Blanc, T. V., 1985: Variation of bulk-derived surface flux, stability and roughness results due to the use of different transfer coefficient schemes. *J. Phys. Oceanogr.*, **15**, 650–669.
- , 1986: The effect of inaccuracies in weather ship data on bulk-derived estimates of flux, stability, and sea-surface roughness. *J. Atmos. Oceanic Technol.*, **3**, 12–26.
- Brohan, P., N. A. Rayner, M. Vanicek, S. F. B. Tett, and E. C. Kent, 2003: Combined estimates of uncertainties in gridded marine temperature fields due to measurement errors and under-sampling of variability. *Proc. Second JCOMM Workshop on Advances in Marine Climatology (CLIMAR-II)*, Brussels, Belgium, JCOMM, CD-ROM, WMO/TD-No. 1199.
- Brunke, M. A., C. W. Fairall, X. Zeng, L. Eymard, and J. A. Curry, 2003: Which bulk aerodynamic algorithms are least problematic in computing ocean surface turbulent fluxes? *J. Climate*, **16**, 619–635.
- Cayan, D., 1992: Variability of latent and sensible heat fluxes estimated using bulk formulae. *Atmos.–Ocean*, **30**, 1–42.
- Chang, E. K. M., 2005: Effects of secular changes in frequency of observations and observational errors on monthly mean MSLP summary statistics derived from ICOADS. *J. Climate*, **18**, 3623–3633.
- Chou, S.-H., E. Nelkin, J. Ardizzone, R. M. Atlas, and C.-L. Shie, 2003: Surface turbulent heat and momentum fluxes over global oceans based on the Goddard Satellite Retrievals, version 2 (GSSTF2). *J. Climate*, **16**, 3256–3273.
- Clark, N. E., L. Eber, R. M. Laurs, J. A. Renner, and J. F. T. Sour, 1974: Heat exchange between ocean and atmosphere in the eastern North Pacific for 1961–71. NOAA Tech. Rep. NMFS SSFR-682, 108 pp.
- da Silva, A. M., C. C. Young, and S. Levitus, 1994: Anomalies of directly observed quantities. Vol. 2, *Atlas of Surface Marine Data*, NOAA Atlas NESDIS 2, 419 pp.
- Dobson, F., and S. D. Smith, 1988: Bulk model of solar radiation at sea. *Quart. J. Roy. Meteor. Soc.*, **114**, 165–182.
- Duchon, C. E., 1979: Lanczos filtering in one and two dimensions. *J. Appl. Meteor.*, **18**, 1016–1022.
- Eymard, L., and Coauthors, 1999: Surface fluxes in the North Atlantic current during CATCH/FASTEX. *Quart. J. Roy. Meteor. Soc.*, **125**, 3563–3599.
- Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson, 2003: Bulk parameterization of air–sea fluxes: Updates and verification for the COARE algorithm. *J. Climate*, **16**, 571–591.
- Fung, I. Y., D. E. Harrison, and A. A. Lacis, 1984: On the variability of the net long-wave radiation at the ocean surface. *Rev. Geophys.*, **22**, 177–193.
- Gibson, J. K., P. Kallberg, A. Uppala, S. Hernandez, A. Nomura, and A. Serrano, 1999: ERA-15 description. Version 2, ECMWF Reanalysis Project Report Series 1, ECMWF, Reading, United Kingdom, 77 pp.
- Girdiuk, G. V., T. V. Kirillova, and S. P. Malevsky, 1985: Cloudiness influence on the oceanic albedo. *Meteor. Hydrol.*, **12**, 63–69.
- Gleckler, P., and B. Weare, 1997: Uncertainties in global ocean surface heat flux climatologies derived from ship observations. *J. Climate*, **10**, 2764–2781.
- Gulev, S. K., 1994: Influence of space–time averaging on the ocean–atmosphere exchange estimates in the North Atlantic midlatitudes. *J. Phys. Oceanogr.*, **24**, 1236–1255.
- , 1995: Long-term variability of sea–air heat transfer in the North Atlantic Ocean. *Int. J. Climatol.*, **15**, 825–852.
- , 1999: Comparison of COADS release 1a winds with instrumental measurements in the Northwest Atlantic. *J. Atmos. Oceanic Technol.*, **16**, 133–145.
- , T. Jung, and E. Ruprecht, 2002: Climatology and interannual variability in the intensity of synoptic-scale processes in the North Atlantic from the NCEP–NCAR reanalysis data. *J. Climate*, **15**, 809–828.
- , V. Grigorieva, K. M. Selemenov, and O. Zolina, 2003a: Evaluation of surface winds and waves from voluntary observing ship data. *Advances in Applications of Marine Climatology, WMO Guide to the Application in Marine Climatology*, WMO, Geneva, Switzerland, 53–67.
- , —, A. Sterl, and D. Woolf, 2003b: Global climatology of ocean waves from the VOS data. *J. Geophys. Res.*, **108**, 3236, doi:10.1029/2002JC001437.
- , T. Jung, and E. Ruprecht, 2007: Estimation of the impact of sampling errors in the VOS observations on air–sea fluxes. Part II: Impact on trends and interannual variability. *J. Climate*, **20**, 302–315.
- Hasse, L., and S. D. Smith, 1997: Local sea surface wind, wind stress, and sensible and latent heat fluxes. *J. Climate*, **10**, 2711–2724.
- Josey, S. A., 2001: A comparison of ECMWF, NCEP–NCAR, and SOC surface heat fluxes with moored buoy measurements in the subduction region of the northeast Atlantic. *J. Climate*, **14**, 1780–1789.
- , D. Oakley, and R. W. Pascal, 1997: On estimating the atmospheric longwave flux at the ocean surface from ship meteorological reports. *J. Geophys. Res.*, **102** (C13), 27 961–27 972.
- , E. C. Kent, and P. K. Taylor, 1999: New insights into the ocean heat budget closure problem from analysis of the SOC air–sea flux climatology. *J. Climate*, **12**, 2856–2880.
- , R. W. Pascal, P. K. Taylor, and M. J. Yelland, 2003: A new formula for determining the atmospheric longwave flux at the ocean surface at mid-high latitudes. *J. Geophys. Res.*, **108**, 3108, doi:10.1029/2002JC001418.
- Jung, T., 2005: Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **131**, 1045–1073.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajaopalan, 1998: Analyses of global sea

- surface temperatures 1856–1991. *J. Geophys. Res.*, **103**, 18 567–18 589.
- , Y. Kushnir, and M. A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854–1992. *J. Climate*, **13**, 2987–3002.
- , M. A. Cane, and Y. Kushnir, 2003: Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects. *Advances in the Applications of Marine Climatology*, World Meteorological Organization, WMO Tech. Doc. 1081, 199–216.
- Kent, E. C., and P. K. Taylor, 1995: A comparison of sensible and latent heat flux estimates for the North Atlantic Ocean. *J. Phys. Oceanogr.*, **25**, 1530–1549.
- , and D. I. Berry, 2005: Quantifying random measurement errors in voluntary observing ship meteorological observations. *Int. J. Climatol.*, **25**, 843–852.
- , P. K. Taylor, B. S. Truscott, and J. S. Hopkins, 1993a: The accuracy of voluntary observing ships' meteorological observations—Results of the VSOP-NA. *J. Atmos. Oceanic Technol.*, **10**, 591–608.
- , R. J. Tiddy, and P. K. Taylor, 1993b: Correction of marine air temperature observations for solar radiation effects. *J. Atmos. Oceanic Technol.*, **10**, 900–906.
- , P. Challenor, and P. K. Taylor, 1999: A statistical determination of the random errors present in VOS meteorological reports. *J. Atmos. Oceanic Technol.*, **16**, 905–914.
- , P. K. Taylor, and P. Challenor, 2000: The effect of successive correction on variability estimates for climatological datasets. *J. Climate*, **13**, 1845–1857.
- Kidson, J. W., and K. E. Trenberth, 1988: Effects of missing data on estimates of monthly mean general circulation statistics. *J. Climate*, **1**, 1261–1275.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- Kubota, M., A. Kano, H. Muramatsu, and H. Tomita, 2003: Intercomparison of various surface latent heat flux fields. *J. Climate*, **16**, 670–678.
- LabSea Group, 1998: The Labrador Sea Deep Convection Experiment. *Bull. Amer. Meteor. Soc.*, **79**, 2033–2058.
- Large, W. G., and S. Pond, 1982: Sensible and latent heat fluxes over the ocean. *J. Phys. Oceanogr.*, **12**, 463–482.
- Legler, D., 1991: Errors in five-day mean surface wind and temperature conditions due to inadequate sampling. *J. Atmos. Oceanic Technol.*, **8**, 705–712.
- Lind, R. J., K. B. Katsaros, and M. Gube, 1984: Radiation budget components and their parameterization in JASIN. *Quart. J. Roy. Meteor. Soc.*, **110**, 1061–1071.
- Lindau, R., 2000: *Climate Atlas of the Atlantic Ocean Derived from the Comprehensive Ocean-Atmosphere Data Set*. Springer-Verlag, 488 pp.
- Lo, A. K., and G. A. McBean, 1978: On the relative errors in methods of flux calculations. *J. Appl. Meteor.*, **17**, 1704–1711.
- Malevsky, S. P., G. V. Girdiuk, and B. Egorov, 1992: *Radiation Balance of the Ocean Surface*. Hydrometeoizdat, 148 pp.
- Moore, G. W. K., and I. A. Renfrew, 2002: An assessment of the surface turbulent heat fluxes from the NCEP–NCAR reanalysis over the western boundary currents. *J. Climate*, **15**, 2020–2037.
- NCAR, 1999: Comprehensive Ocean-Atmosphere Data Set (COADS): Long marine reports/fixed marine reports (LMR6/LMRF.6). Release 1a Documentation. [Available online at <http://icoads.noaa.gov/e-doc/lmrf/>]
- Niekamp, K., 1992: Untersuchung zur Gute der Parametrisierung von Malevsky-Malevich zur Bestimmung der solaren Einstrahlung an der Ozeanoberfläche. (Investigation of the skills of Malevsky-Malevich parameterization of short-wave radiation at the ocean surface). M.Sc. thesis, Department of Marine Meteorology, Institut für Meereskunde, Kiel, Germany, 108 pp.
- Oberhuber, J. M., 1988: An atlas based on the COADS data set: The budgets of heat, buoyancy and turbulent kinetic energy at the surface of the global ocean. MPI Rep. 15, 19 pp.
- Oort, A., 1978: Adequacy of the rawinsonde network for global circulation studies tested through numerical model output. *Mon. Wea. Rev.*, **106**, 174–195.
- Payne, R. E., 1972: Albedo at the sea surface. *J. Atmos. Sci.*, **29**, 959–970.
- Reed, R. K., 1977: On estimating insolation over the ocean. *J. Phys. Oceanogr.*, **7**, 482–485.
- Renfrew, J., G. W. K. Moore, P. S. Guest, and K. Bumke, 2002: A comparison of surface-layer, surface heat flux and surface momentum flux observations over the Labrador Sea with ECMWF analyses and NCEP reanalyses. *J. Phys. Oceanogr.*, **32**, 383–400.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climatol.*, **7**, 929–948.
- Rouault, M., C. J. C. Reason, J. R. E. Lutjeharms, and A. C. M. Beljaars, 2003: Underestimation of latent and sensible heat fluxes above the Agulhas Current in NCEP and ECMWF analyses. *J. Climate*, **16**, 776–782.
- Schulz, J., J. Meywerk, S. Ewald, and P. Schlüssel, 1997: Evaluation of satellite-derived latent heat fluxes. *J. Climate*, **10**, 2782–2795.
- Smith, S. R., D. M. Legler, and K. V. Verzone, 2001: Quantifying uncertainties in NCEP reanalyses using high-quality research vessel observations. *J. Climate*, **14**, 4062–4072.
- , J. Servain, D. M. Legler, J. N. Stricherz, M. A. Bourassa, and J. J. O'Brien, 2004: In situ-based pseudo-wind stress products for the tropical oceans. *Bull. Amer. Meteor. Soc.*, **85**, 979–994.
- Smith, T. M., and R. W. Reynolds, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate*, **16**, 1495–1510.
- , and —, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate*, **17**, 2466–2477.
- Sterl, A., 2001: On the accuracy of gap-filling algorithms in global surface fields. *Geophys. Res. Lett.*, **28**, 2473–2476.
- , 2004: On the (in)homogeneity of reanalysis products. *J. Climate*, **17**, 3866–3873.
- Taylor, J. R., 1982. *An Introduction to Error Analysis*. University Science Books, 270 pp.
- The WGASF Group, 2000: Intercomparison and validation of ocean-atmosphere energy flux fields: Report of the Joint WCRP/SCOR WG on Air-Sea Fluxes, P. K. Taylor, Ed., WMO, 305 pp. [Available online at <http://www.soc.soton.ac.uk/JRD/MET/WGASF/>]
- Uppala, S., and Coauthors, 2005: The ERA-40 reanalysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Weare, B. C., 1989: Uncertainties in estimates of surface heat fluxes derived from marine reports over the tropical and subtropical oceans. *Tellus*, **41A**, 357–370.

- , and P. T. Strub, 1981: The significance of sampling biases on calculated monthly mean oceanic surface heat fluxes. *Tellus*, **33**, 211–224.
- Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS release 2 data and metadata enhancements for improvements of marine surface flux fields. *Phys. Chem. Earth*, **23**, 517–526.
- Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.*, **25**, 823–842.
- Yu, L., R. A. Weller, and B. Sun, 2004a: Improving latent and sensible heat flux estimates for the Atlantic Ocean (1988–99) by a synthesis approach. *J. Climate*, **17**, 373–393.
- , —, and —, 2004b: Mean and variability of the WHOI daily latent and sensible heat fluxes at in situ flux measurement sites in the Atlantic Ocean. *J. Climate*, **17**, 2096–2118.
- Zeng, X., M. Zhao, and R. Dickinson, 1998: Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data. *J. Climate*, **11**, 2628–2644.
- Zhang, H.-M., J. J. Bates, and R. W. Reynolds, 2006: Advances in higher resolution Global Ocean Observing System: Sea surface wind speed perspective. Preprints, *14th Conf. on Interaction of the Sea and Atmosphere*, Atlanta, GA, Amer. Meteor. Soc., CD-ROM, 6.5.
- Zolina, O., and S. K. Gulev, 2003: Synoptic variability of ocean–atmosphere turbulent fluxes associated with atmospheric cyclones. *J. Climate*, **16**, 3023–3041.